The Dissertation Committee for Han Ren Certifies that this is the approved version of the following dissertation:

Academic Achievement Outcomes of Latino English-Language Learners in Texas: A Longitudinal Analysis

Committee:

Stephanie W. Cawthon, Supervisor

Keenan A. Pituch

Marilla D. Svinicki

Timothy Z. Keith

Molly A. Lopez

Academic Achievement Outcomes of Latino English-Language Learners in Texas: A Longitudinal Analysis

by

Han Ren, B.A., M.A., M.Ed.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2015

Dedication

I dedicate this dissertation to my father, Dr. Shenyang Guo. This degree and the work it represents would not be possible without your tireless support, unconditional love, and model of excellence throughout my life.

Acknowledgements

I would like to acknowledge those whose support and contributions have made this work possible. My dissertation chair and graduate school adviser, Dr. Stephanie Cawthon provided essential feedback, support, and mentorship throughout these past seven years. This dissertation topic was born from an idea sparked through working in her Accommodations lab during my first semester at UT. With her help, I was able to nurture it into fruition. I owe much of my overall development as a researcher to Dr. Molly Lopez, who took me under her wing at the Center for Social Work Research during my first years as a graduate student. I would like to express my sincerest appreciation for the other members of my committee, Drs. Keenan Pituch, Timothy Keith, and Marilla Svinicki, for their support, expertise, and fantastic suggestions during this process.

I am truly in debt and thankful to my fellow Texas ERC researcher, Dr. Amy Wiseman. Her statistical expertise and familiarity with this vast database provided me with a roadmap for my analyses. To the essential members of the Texas ERC, Dr. Celeste Alexander, Cindy Corn, and members of the board, thank you for taking a chance on my proposal and allowing me access to this valuable data.

I owe my sanity to the patience, love, good humor, and grace of my family and friends. To my husband, Justin, you have supported me in every way since the beginning, one bite at a time. Thank you for believing in my dreams. To my darling daughter, Olive, you are my light and inspiration. You motivate me to be better in all facets of life, especially in finding balance and expressing joy. To my parents, colleagues, and friends who kept me grounded and focused throughout this journey, I appreciate you beyond measure.

Academic Achievement Outcomes of Latino English-Language Learners in Texas: A Longitudinal Analysis

Han Ren, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Stephanie Cawthon

Texas has one of the highest populations of English Language Learners (ELLs) in the U.S., with a complex system for ELL identification, program placement, and high-stakes assessment. Spanish-speaking Latino ELLs represent a large proportion of this population in this state. The long-term academic achievement of ELLs identified in elementary grades and educated in different program placements is not well known. Prior research presents support for Bilingual Education models as most promising for future student achievement. Using strict sampling criteria and analysis of variance (ANOVA) procedures, this study compared student achievement during secondary years, as measured by high-stakes assessments, of demographically similar Latino former-ELLs, long-term ELLs, and non-ELLs who received a variety of language service programming during elementary school.

This study examined data from a cohort of 18,188 students enrolled in all 10 major urban school districts in Texas from 3rd to 9th grades, between the 2003-2004 and 2009-2010 school years. Math and reading high-stakes test scores during students' 7th and 9th grade years were used as outcome variables. Key findings show that ELLs who did

not receive any language programming have significantly higher achievement outcomes than ELLs who received English as a Second Language (ESL), Bilingual Education, or equal years of ESL and Bilingual. These students performed on par with their non-ELL peers in both math and reading. Results also show achievement differences between former-ELLs who were able to achieve English language proficiency by the start of middle school, and long-term ELLs who remained ELL through secondary education. Former-ELLs had favorable outcomes, with math achievement scores that were higher than non-ELLs, whereas long-term ELLs performed significantly lower than both former-ELLs and non-ELLs at all study points. Interactions between programming received and ELL-type, as well as the effects of immigration status and district of enrollment were explored.

Results were unexpected and highlight the need for further longitudinal research on existing ELL populations across the state of Texas. Implications of this study support current educational policies that shift away from prolonged Bilingual Education toward ESL and mainstream-English only programming for ELLs. Additional policy implications and directions for future research are discussed.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

English Language Learners (ELLs) are among the fastest growing student populations in the U.S. public school system, with Texas ranking second highest of all states for the number of students attaining English proficiency (U.S. Department of Education, 2006). The growth of ELLs has been increasing at a faster rate compared to overall student enrollment in recent years. Between the 1995-96 and 2005-06 school years, total enrollment in the state of Texas increased 13.6% while ELL enrollment increased 34.0% (National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, 2007). Among this diverse group of ELLs, Latino students represent the fastest rising group, comprising one out of every five U.S. public school students in 2008 (Fry & Gonzales, 2008). As a state with one of the highest numbers of ELLs, Texas bears strong responsibility to establish effective policies in the identification, education, and assessment of these students. As ELL populations continue to grow within public school systems, other states will look to Texas as an exemplar for ELL education models. The need to identify, educate and assess the growing population of ELL students across the nation presents a formidable challenge for researchers, policy makers and educators alike.

The issue of ELL education has been a prominent research and policy topic both historically and more recently within the context of the national education reform movement. The history of bilingual education is long and contentious, marked by litigation by minority groups in the pursuit of educational equality (San Miguel, 2004; Valencia, 2008). Contemporary bilingual education policy was established in the 1960's,

with the passage of the Bilingual Education Act of 1968. This act coupled with the Equal Educational Opportunity Act (EEOA) of 1974 paved the way for federal legislation enforcing equitable and differentiated education for language minority students (Stewner-Manzanares, 1988; Valencia, 2008). Supporters of bilingual education consider language equality to be a civil rights issue, with appropriate educational access for ELLs as a vital component for achieving a multicultural America (San Miguel, 2004). Conversely, bilingual education has been met with much opposition from English-only education proponents, who fear that language education will divide America into a bicultural and bilingual society, undermining the country's foundational Anglo-Protestant roots (Hempel, Dowling, Boardman, & Ellison, 2012; Huntington, 2004). This sociopolitical debate continues to play out in recent educational reform policies.

On a policy level, the No Child Left Behind Act of 2001 (NCLB) essentially eliminated the federal provision of bilingual education by replacing it with English-only legislation in order to meet the linguistic needs of ELLs (San Miguel, 2004). While local and state educational agencies may still offer bilingual programming if perceived to be appropriate, federal funding incentives place emphasis on ELLs attaining English proficiency as quickly as possible, promoting the English-only educational agenda. Additionally, the United States Department of Education allows each state to devise their own rules and regulations surrounding the identification, education, and assessment of ELLs, resulting in inconsistent laws and practices across states. NCLB (2002) requires all students to participate in high-stakes testing as part of a federal accountability system, whereas ELL data were formerly excluded from school-wide score reports. This

requirement places the onus on schools to properly educate ELLs and ensure they are receiving an equitable education as their Native English Speaker (NES) peers, as federal funding are now contingent on all student test scores in English.

While federal policy aims to ensure all students receive an equitable education, the amalgamation of these policies in practice has yielded dismal achievement outcomes for ELLs. The achievement gap between ELLs and their NES peers is well established (U.S. Department of Education, 2009; 2010; 2012a; 2012b; 2012c), as illustrated through higher dropout rates, lower academic achievement, and an overrepresentation of ELLs in special education programs compared to their NES peers. There is a significantly lower proportion of ELLs who pass statewide high-stakes tests under NCLB (2002) compared to their NES peers, especially at the secondary level (Smith, 2010; TEA, 2008). This achievement gap has sparked researchers to examine the root causes of this educational inequity and track the progress of current and former ELLs as they navigate school systems, in order to develop and propose viable policy solutions.

At the center of this research are the development and evaluation of effective instructional programming for ELLs. Taking into consideration culture, context, policy and second language acquisition theory, scholars aim to identify ELL classroom models with the most promising long-term achievement outcomes. Instructional programming for ELLs is in a state of constant evolution, with the identification of six or more program models, ranging from dual-language bilingual education to English-only structured immersion classrooms (Thomas & Collier, 2002; Genesee, 1999). The availability of different program models varies significantly among states and school districts, resulting

in inconsistent provision of services. However, the most prevalent language instruction program models can be reduced to three categories: mainstream English-only, Bilingual Education, and English as a Second Language (ESL) (Genesee, 1999). The establishment of the most effective instructional model is a hotly contested research topic, especially given the constraints of shifting educational policies (Thomas & Collier, 2002). While some scholars identify bilingual education as producing the most promising achievement outcomes for ELLs (August & Shanahan, 2006; Genesee, Lindholm-Leary, Saunders, & Christian, 2006; Goldenberg, 2008; Mora, Wink, & Wink, 2001; Thomas & Collier, 1997; 2002), other studies fail to find clear advantages between language instruction models (Galvan-Luis, 2010; Nakamoto, Lindsey, & Manis, 2012; Slavin, Madden, Calderon, Chamberlain, & Hennessy, 2011). There is a lack of consensus on what constitutes best practice for the identification, education, and assessment of ELLs (Abedi, 2004; August & Hakuta, 1997). However, based on the achievement gap, it is clear that current educational systems are not adequately meeting their academic needs.

Research on current ELLs has focused on achievement in high school, examining the impacts of accountability, policy, and language on dropout rates and academic achievement on Spanish-speaking students primarily from low-socioeconomic status (SES) backgrounds within urban school districts (McNeil, Coppola & Vasquez-Heilig, 2008; Vasquez-Heilig & Darling-Hammond, 2008). Through this research, a group of unique ELLs has emerged termed "long-term ELLs," who require five or more years to attain English proficiency (Menken & Kleyn, 2009; Olsen, 2010). This subset of ELLs faces exceptional challenges, with poor achievement outcomes. They are frequently

adolescents who appear orally bilingual, yet lack academic literacy skills in either their native language or English (Menken & Kleyn, 2009; Olsen, 2010). These students are included in the current study in hopes of understanding some of the factors that allow some ELLs to achieve English proficiency, while others remain perennially stuck in ELL classification.

Although there is considerable evidence that current ELLs have significantly lower academic achievement than NES students, the literature on long-term achievement outcomes for students who have been reclassified from ELL to English language proficient is just beginning to emerge. A growing body of recent research supports positive long-term achievement outcomes for these former-ELLs (Ardasheva, Tretter, & Kinny, 2012; Flores, Batalova, & Fix, 2012; Halle, Hair, Wandnar, McNamara, & Chien 2012; Kim, Herman, & National Center for Research on Evaluation, 2009; Kim, Herman, & National Center for Research on Evaluation, 2012). Factors that contribute to favorable former-ELL achievement outcomes include maintaining grade-level academic proficiency, three years of language programming, and earlier age of reclassification (Flores, Batalova, & Fix, 2012; Halle et al., 2012).

The literature comparing former-ELLs who receive differential instruction is still lacking. The research that does compare student achievement outcomes post-reclassification from different program models (de Jong, 2004; Kim & Herman, 2009; Thomas & Collier, 2002) rarely follows students past middle school. Most of these studies also utilized data collected before the implementation of NCLB (2002), which has significantly altered the academic environment, as ELLs' test scores now contribute to

federal funding decisions. Furthermore, the ELL populations featured in prior studies come from various backgrounds, differing in native language, SES, state of enrollment, and geographic setting (rural versus urban). Thus it is difficult to draw conclusions on the long-term academic trajectories of these former-ELLs with diverse backgrounds, and there may be a cohort effect as a result of changing educational policies. Further understanding of a cohort of ELLs as they mature and progress through the educational system will provide a clearer understanding about the academic trajectories of ELLs, and the relative effectiveness of various elementary language instruction models on future achievement.

This study retrospectively examines student data from the 10 major urban Texas school districts as of 2009-2010, to investigate the effects of available classroom program models on future academic achievement for two groups of demographically similar ELLs identified in elementary school. The achievement trajectories of long-term ELLs and former-ELLs compared to their non-ELL peers will be compared. This cohort of students is an important population of study due to the lack of demographically controlled, longitudinal analysis on the long-term academic achievement of ELLs in previous research. Through examining this data, one might be able to infer the effectiveness of recent available academic models for preparing ELLs to exit language programming and enter mainstream education. The contributing factors and achievement differences between long-term ELLs and former-ELLs will be reviewed in order to assess why some ELLs are able to attain English proficiency while others struggle.

In particular, this study will evaluate student high-stakes test score differences between ESL and bilingual education program models, as well as in comparison to English-only mainstream education. Due to the variety and inconsistency of language programming availability, the conglomerate categories of ESL and bilingual education will be used to maximize statistical power, and to simplify analyses. To control for the effects of diverse backgrounds, only Spanish-speaking ELLs from low-SES families will be included in the target groups. The comparison group will be comprised of low-SES, non-ELL, Latino students. Using rigorous sampling methods and univariate analysis of variance, this study will compare the high-stakes test scores of a cohort of ELLs enrolled in two language program models and mainstream English-only instruction at two points in their education, in 7th and 9th grade, in order to compare their relative implications for student achievement.

Chapter Two: Review of the Literature

According to the U.S. Department of Education's National Center for Educational Statistics (2012), the number of school-aged ELLs rose from 4.7 million to 11.2 million between 1980 and 2009. These numbers continue to rise, especially in Texas, where over 838,000 ELLs were in enrolled during the 2011-2012 school year, compared to 775,000 ELLs four years ago, growing from 16% to 17% of the total student population (Texas Education Agency, 2009; 2012a). As ELLs make up an increasing proportion of all students, the questions of how to adequately educate, assess, and track this population have become common interests for educators, researchers, and policy makers alike. Because of the mandates of the No Child Left Behind Act of 2001 (NCLB) that require ELLs to participate in high-stakes assessments tied to state funding of public education, the stakes for understanding ELL achievement are higher than ever. This chapter explores the growing body of research on how ELLs are best served, including topics such as language acquisition theory, classroom models of instruction, program duration, and achievement outcomes.

This literature review begins with a discussion on the delineation between LEP and ELL as two commonly used terms to describe the same student population. The current policies that guide how schools identify and track ELL students are also explained. Next, I summarize the history of bilingual education in the U.S. to provide a background context and introduce the socio-political dialogue around current educational policies on the identification, education, and assessment of ELLs. The achievement gap of the differential academic performance between Latino students, many of whom are

ELLs, and their White counterparts will be highlighted to explain the practical importance of research in this area. The theoretical constructs behind paradigms of language education and second language acquisition will be explored to provide a framework for the various classroom models of instruction available for ELLs. I then incorporate these language perspectives to describe available educational program models in Texas and their effects on ELLs. Finally, the last sections examine the literature on the achievement patterns of long-term and former-ELLs, and describe the current condition of educational policies in Texas to provide rationale for this study.

A Note on Terminology

There are many terms used to describe students who are not proficient in English, including English Language Learner (ELL), Limited-English proficiency (LEP), language minority (LM), second language learners (SLL), and non-English speakers (U.S. Department of Education, 2013). These terminologies are similar in meaning as each refers to students who are acquiring English as a second language before achieving fluent oral and academic English language proficiency. In this document, I use both “ELL” and “LEP” to describe this student population, however the term used depends on context. "ELL" is the newer and preferred term used in schools and educational research, as it labels the learner positively, whereas "LEP" ascribes a deficiency to the learner (National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, 2009). The term "LEP" is still used today in federal and state documents such as NCLB (2002), as well as in a historical context. Terminology varies across states and policy contexts. Officially in Texas, “ELL” and “LEP” are used

interchangeably (Texas Education Code §89.1203). For the purposes of this document, "ELL" will be the predominantly used term for describing this population, except when referring directly to state, federal, or historical policy contexts, which continue to refer to these students as "LEP."

The interchangeable use of “ELL” and “LEP” in Texas terminology is a relatively recent development (Texas Education Code §89.1203). Prior to the 2011 school year, Texas policy used the term “LEP” as an eligibility status describing recent immigrants within their first three years of high-stakes testing who were afforded language accommodations or a Spanish-version of the statewide assessment (Texas Education Code §101.1007, 2008). The term “ELL” generally described students in the process of attaining English proficiency, regardless of whether language services were provided (Cawthon, 2010). For the purposes of clarity, this document will continue to refer to Texas ELLs eligible for language provisions on high-stakes assessments during their first three years of U.S. schooling as “LEP,” as the timeframe of this study uses data collected before the Texas terminology shift in 2011. The term “ELL” is used in most contexts to broadly define all students who are not yet English proficient.

ELL Identification and Policy

The No Child Left Behind Act (NCLB) of 2001 (P.L. 107-100) was the first major legislative reform of the Elementary and Secondary Education Act (ESEA) of 1965 (P.L. 89-10), with an emphasis on accountability for all learners. Under NCLB (2002), statewide accountability systems covering all public schools were implemented to ensure each school made Adequate Yearly Progress (AYP), a requirement tied to federal

funding. Implications for ELLs were vast, as their academic performance and test scores were evaluated on a national scale for the first time in history. NCLB (2002) provided basic guidelines for the identification and education of ELLs, while allowing individual states the flexibility to define how these students are assessed and tracked. This section focuses on identification and assessment policy, while available educational programming for these students will be discussed later.

Title IX, Part A, Sec 9101 (25) of NCLB identifies LEP students as meeting the following criteria: between ages 3-21, enrolled in elementary or secondary school, not born in the U.S. or their native language is not English, Native American or Alaska Native, from an environment where a non-English language has had a significant impact on their English proficiency, and have difficulty with speaking, reading, writing, or understanding English that may prevent the student from meeting state proficiency levels of achievement or succeeding in English-only classrooms. Considering this broad categorization of ELLs, in practice, this determination is most often based on the students' Home Language Survey (HLS). This survey was originally recommended in the *Lau Remedies* of 1975 as the primary source of information for determining students' LEP eligibility, and is still widely used today.

The HLS in Texas is given to all students and asks the questions: "What language is spoken in your home most of the time?" and "What language does your child speak most of the time?" (Texas Education Code, §89.1215). Students with a home language that is not English, or English in addition to a non-English language, are assessed on their current level of oral English proficiency. Students in 2nd grade and above are also given a

norm-referenced standardized achievement test. Students in prekindergarten through 1st grade meet ELL criteria based on oral English proficiency alone, whereas older students must also score below the 40th percentile on the standardized achievement test to receive ELL designation (Texas Education Code §89.1225). Students who meet ELL criteria must be offered access to appropriate language instruction programming and language accommodations on statewide assessments.

After students are initially identified as ELL, federal and state guidelines determine how these students are educated, assessed, and tracked. On the federal level, Title III of NCLB (2002) details language instruction for LEP and immigrant students. According to Sec. 200.6(b)(4)(iv) of NCLB (2002), recent immigrant students are considered LEP for their first 12 months of schooling in the U.S., and exempt from statewide high-stakes testing. After this initial grace period, LEP students are required to participate in statewide testing, with scores contributing to each school's AYP accountability. However, there remains flexibility in how states assess these students.

In Texas, special language provisions are available for the first three years of immigrant ELLs' U.S. schooling. These "LEP eligible" services include Spanish-versions of assessments for students in grades 3-5, and linguistically accommodated English assessments for students older than 5th grade and those whose native language is not Spanish (Texas Education Code §101.005). After three years, these students are no longer considered LEP eligible, regardless of their English proficiency. Therefore, many Texas students continue to be ELLs past the allowable three-year LEP designation period (Cawthon, 2010). Parents have the option of accepting language instruction support, or

waiving this right and enrolling their children in mainstream English monolingual programming (Texas Education Code, rule §89.1240).

In sum, there are three major distinct but related categories to consider in the identification, education, and assessment policies of ELLs in Texas: ELL, LEP-eligible, and instructional programming. First, all ELLs are defined as students in the process of learning English who are not yet proficient enough to be successfully educated in mainstream English-only classrooms (NCLB, 2002). A subset of ELLs is LEP-eligible new immigrants who are afforded language provisions during their first three years of statewide testing. These students are eligible for exemption on high-stakes testing during their first year (NCLB, 2002), and Spanish or language-accommodated English versions of statewide assessments for an additional two years (Texas Education Code §101.005). After three years, these students are assessed using the same tests as their Native-English Speaking (NES) peers, regardless of English proficiency. Until redesignation as English proficient, and for two years after, ELL and former-LEP students are tracked with their test scores falling under the ELL subset for the purposes of AYP accountability (Sec. 3121, Title III, NCLB, 2002). Once ELLs are identified, they are eligible to receive special instructional programming to help facilitate academic English acquisition under Title III of NCLB (2002). However, the availability, determination, and effectiveness of different language-instruction models of education are hotly contested from both research and socio-political perspectives. There is not one uniform answer to this complex educational question, resulting in diverse instructional programming and outcomes for ELLs. The history and research contexts of language programming are explored below.

History of Bilingual Education

Bilingual education in the U.S. has a long and contentious past, marked by litigation and opposing sociopolitical and cultural viewpoints rooted in immigration policy and systemic inequity (San Miguel, 2004; Valencia, 2008). One cannot enter the bilingual education discussion without first acknowledging the deeply embedded contextual factors that have shaped its evolution, especially in a state like Texas, with its geological proximity to Mexico and large ELL population. Before delving into the history of bilingual education, I will give context to this debate by exploring the theoretical underpinnings and arguments of stakeholders on both sides of the bilingual education discussion.

Pro-bilingual education scholars often speak from the perspective of Critical Race Theory, which recognizes the intersection between race and inequality, regards racism as deeply endemic in American society, challenges dominant ideology, and grounds its conceptual framework in the unique experiences of people of color (Ladson-Billings & Tate, 1995; Valencia, 2008). Perspectives toward bilingual and Latino education center around themes of "plight" and "struggle," highlighting the historical oppression experienced by ELLs in school systems, and the continued fight for equitable education for all learners (San Miguel, 2004; Valencia, 2008). The issue of bilingual education represents only one of many movements toward equality within the dominant American culture, whose historic systems inherently place people of color at a disadvantage (Ladson-Billings & Tate, 1995).

On the other side of the debate are champions for English-only education models,

proponents of which often hold Anglo-Protestant centered perspectives of the "American Dream," focused on achieving a nation of unified culture and language (Huntington, 2004). These views may be explained by Racial Threat Theory, which posits that as minority populations increase, the dominant culture responds through increased segregation, symbolic racism, and perceived threat to the majority's private lives and the status quo (Kinder & Sears, 1981; Hempel, Dowling, Boardman & Ellison, 2012). Huntington (2004) attributes immense immigration, irreconcilable cultural differences, and lack of assimilation by Latino-Americans as major contributors to the creation of a divided United States with two distinct languages and cultures. Hempel et al. (2012) explored the individual and contextual factors influencing opposition to bilingual education in Texas. Consistent with Racial Threat Theory, researchers found that politically conservative, older Americans with lower SES and education residing in high-Latino areas with continued substantial Latino growth were most likely to support English-only education programs (Hempel et al., 2012). The sociopolitical discussion around bilingual education continues as evidenced by ongoing policy changes in the national education reform movement. Advocates of bilingual education strive for multicultural equality in all domains, while opponents aim to maintain the established Anglo-Protestant status quo. Now that the cultural context has been established, let us examine the historical pathways that have allowed bilingual education to evolve to its current state.

Although the origins of bilingual education in the U.S. extend back to the 1800's in recognizing the language rights of Native Americans (Valencia, 2008), it was not until

50 years ago that modern bilingual education policy began to solidify. The 1960's served as a significant time period in the development of contemporary bilingual education. During this decade, the combination of the Civil Rights Movement, federal social legislation, and research findings on bilingualism, paved the way for the passage of the Bilingual Education Act of 1968 (Stewner-Manzanares, 1988; Valencia, 2008). Although the Civil Rights Movement originally began as the African American struggle for equality, other issues of discrimination were soon brought to light during this period of heated social climate. Language scholars argued that discrimination extends beyond race, and encompasses issues of national origin, gender, religion, language and culture (San Miguel, 2004). This eventually led to federal social legislation enforcing an equitable education for LEP students. Title VI (42 U.S.C. § 2000d) of the 1964 Civil Rights Act (P.L. 88-352) prohibits discrimination on the basis of race, color and national origin in programs receiving federal funding, including public education. Another critical federal social legislation was The Elementary and Secondary Education Act (ESEA) of 1965 (P.L. 89-10), which was reauthorized in 2001 as the No Child Left Behind Act (NCLB). Under Title I (20 U.S.C. §241a) of ESEA, children of low-income families were afforded funding for compensatory education, which included the ELL children of immigrants. This act focused educational attention on meeting the linguistic and cultural needs of low-income ELL students.

Research findings in the 1960's on bilingualism provided empirical evidence supporting the value of bilingual education. Prior to this decade, the misconception that bilingualism had detrimental effects on children's cognitive development was a widely

held belief by both educators and researchers. However, in 1962, Peal and Lambert conducted their landmark study, which demonstrated that balanced bilingual children (French and English speaking) have higher intellectual abilities compared with their monolingual peers of a similar background. This study combined with the Civil Rights movement and social legislation helped pave the way for garnering support for bilingual education. In 1968, ESEA was amended to become the first Bilingual Education Act. This legislation served to correct civil rights violations, and established the right of ethnic minorities to seek differentiated education. Additionally, this act provided funding for resources for diverse educational programming, teacher training, parent involvement projects, and the development and dissemination of content material (Stewner-Manzanares, 1988).

The 1970's continued to bring progressive change to bilingual education, beginning with the case of *Lau v. Nichols* (1974). Prior to this class-action suit, participation in the Bilingual Education Act was voluntary, and the law lacked specific implementation guidelines. In *Lau v. Nichols,* the U.S. Supreme Court ruled that a group of Chinese-speaking minority students in San Francisco were being denied access to an equitable education in public schools because of their limited English skills. This ruling gave birth to the Equal Educational Opportunity Act (EEOA) of 1974, which required specific instructional programs, in ELLs' native language when possible, in order to overcome language barriers for all students. Thus the term "Limited English Proficient" was coined, and all students who received this label were required to receive English language instruction in both state and federally funded public schools. In response to this

ruling, the Office for Civil Rights established the *Lau Remedies* in 1975 to provide suggestions for the identification, placement, and education of ELLs. The *Lau Remedies* remain the primary procedural guidelines used in practice. However, the Bilingual Education Initiative of 1985 allowed local school districts to determine the best method of educating the unique LEP population of students in each area. This act allows each state to determine its own regulations surrounding the identification and education of ELLs. As a result, much variability persists among states in the actual implementation of the *Lau Remedies*, which translates to inconsistent practices and may account for poor student outcomes and the existence of the achievement gap.

The new millennium brought about major shifts in educational policy with profound implications for bilingual education. The No Child Left Behind Act of 2001 requires accountability through high-stakes testing for all students in English only. The Bilingual Education Act is now Title III of NCLB (2002), which essentially repealed bilingual education at a federal level and replaced it with English-only legislation (San Miguel, 2004). Local and state educational agencies may still offer bilingual programming if perceived to be appropriate. However, with federal funding contingent on English-only test scores, more school districts are moving away from bilingual education, in an attempt to increase ELLs' English proficiency as quickly as possible. This makes it more difficult to implement full bilingual programming. Additionally, statewide policy changes such as Proposition 227 in California (California Department of Education, 1999), Proposition 203 in Arizona (Arizona Secretary of State, 2000), and Question 2 in Massachusetts (Massachusetts Department of Education, 2002), allow only

one year of intensive instruction for ELLs to gain academic English proficiency. These mandates further promote the English-only agenda.

Recent contradicting developments in the national educational policy landscape continue to shape how ELLs are educated and assessed. The Common Core State Standards Initiative aims to standardize national learning objectives across content domains for all American students (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). This initiative strives to address structural inconsistencies across states in the provision of educational standards by presenting uniform and rigorous learning objectives spanning all grade levels. Forty-three states have adopted these standards, signifying a shift toward a more national view of education that better aligns with the accountability requirements under NCLB (2002). Meanwhile, the U.S. Department of Education (2015) has invited individual state education agencies to apply for a waiver granting flexibility in implementing specific requirements of NCLB (2002), so long as each state provides rigorous and comprehensive individually developed plans to improve educational outcomes for all students. Currently, 45 states have applied for these ESEA flexibility waivers, and 43 states along with the District of Columbia and Puerto Rico have received approval (U.S. Department of Education, 2015). Texas applied for this waiver in February 2013 (Texas Education Agency, 2013) and was approved on September 30, 2013 (U.S. Department of Education, 2015). The opposing goals of the Common Core State Initiative Standards and the ESEA flexibility waivers demonstrate that educational policy remains polarized on a national level, with simultaneous movement towards both national standardization

and state individualization.

On a state level, recent litigation highlights the movement towards autonomy for state education agencies in establishing ELL policy. In the case of *Horne vs. Flores* (2009), the U.S. Supreme Court overturned the decision of the state court of Arizona, which ruled that the state was violating EEOA by not allocating appropriate funding to cover the costs of ELL instruction. The U.S. Supreme Court ruled in favor of Horne, allowing the state to determine its own requirements regarding the instruction of ELLs. A similar ruling occurred in Texas in 2010, when the 5th U.S. Circuit Court of Appeals overturned a lower court's decision to overhaul the bilingual and ESL programs in Texas schools (Roebuck, 2010). Both of these cases suggest that despite a national movement towards the standardization of education, policy changes that limit individual states' autonomy is more difficult to implement, especially as it relates to funding and the education of ELLs. This polarization may be due in part to the contentious history of bilingual education in the U.S., the dearth of conclusive research demonstrating the superiority of any one educational model, conflicting interests between bilingual education advocates and English-only supporters in regard to educational programming, and the lack of standardization in defining, assessing, and educating ELLs. Regardless of policy, the achievement statistics of ELLs and Latino students clearly demonstrate the existence of a systemic problem.

The Achievement Gap

The "achievement gap" in education refers to the disparity between groups of students on their academic performance. Although an achievement gap may exist

between any two groups of students, it is most commonly used to describe the gap between low-performing African-American and Latino students' academic performance compared to their high-achieving White counterparts (e.g., achievement test scores, U.S. Department of Education, 2010). A similar gap appears between students from low-income backgrounds, compared to their more affluent peers, which may account for differences in achievement between racial groups (Lubienski, 2002). The achievement gap is revealed when comparing student grades, standardized test-scores, dropout rates, course selection, and graduation rates. Given the nature of the current research, our focus will be on the achievement gap between Latino students (many of whom are ELLs) and their White peers.

There are many ways to assess student achievement, and the existence of the achievement gap shows up on each of these measures. Student test scores on standardized assessments are a widely accepted indicator of academic performance, as are retention rates, dropout rates and graduation rates. The National Assessment of Educational Progress (NAEP) measures how well students have learned the formal curriculum, and assesses academic performance across content areas in grades 4, 8, and 12. In 2011, while 91% of White 4th graders performed at or above basic level in math, only 72% of Latino students achieved basic proficiency (U.S. Department of Education, 2012b). Reading results from the same year show a similar pattern, with 78% of White 4th graders achieving basic proficiency, compared with 51% of Latino students (U.S. Department of Education, 2012c). Statistics for ELLs are even more disturbing, with only 31% of fourth grade ELLs performing at or above basic proficiency on the 2011 reading test, compared

to 72% of non-ELL 4th graders (U.S. Department of Education, 2012c).

In Texas, there exists a substantial achievement gap between ELL students and their NES peers, especially during secondary education. The percentage of ELL students passing the Texas Assessment of Knowledge and Skills (TAKS), the state's high-stakes assessment used from 2002-2012, across subjects shows a steep decline after middle school (from 6th to 7th grade), and continues to drop during high school (Smith, 2010; TEA, 2008). Between 2002 and 2007, approximately 70% of ELL 3rd graders passed the TAKS, compared to fewer than 20% of ELL 11th graders (Smith, 2010), a pattern consistent throughout this five-year span. This alarming trend demonstrates a significant problem, as the number of new immigrant ELL students in secondary school continues to grow. Furthermore, the sharp decline in passing rates between elementary and middle school questions the effectiveness of secondary instructional and assessment policies for ELL students. Current assessment policy offers Spanish versions of the state's high-stakes assessment only from grades 3-5 (Texas Legislature Online, 2009), and refuses linguistic accommodations to students with emerging English proficiency after a three-year grade period, which may contribute to their subsequent low performance (Cawthon, 2010).

The achievement gap between White and Latino students is further illustrated when examining high school dropout and graduation rates. In 2007, the total percentage of 16- through 24-year-olds who were not enrolled in high school, and without a high school credential, was 9.8% of males, and 7.7% of females (U.S. Department of Education, 2009). For Latino youth, this percentage rose to 24.7% of males and 18% of

females, representing the racial group with the highest percentage of dropouts. In comparison, only 6% of White males, and 4.5% of White females fit this demographic in 2007. As for graduation rates, 76.2% of White students compared to 57.8% of Latino students graduated high school across the country in the 2003-04 school year (Swanson, 2008). These statistics have improved slightly over time but the pattern remains consistent, with 15.1% of Latino student dropouts in 2010, compared to 5.1% of their White counterparts. The Latino population remains the racial group with the highest percentage of high school dropouts (U.S. Department of Education, 2012a). These trends clearly indicate that Latino students are performing significantly worse than White students on several national measures of student achievement.

While there is not a clear etiology for the achievement gap, educational research suggests several factors that may contribute to this discrepancy. The most obvious and salient of these is language difference. Students who are unable to effectively communicate in the dominant language are clearly at a disadvantage. Through enrollment in language instruction programming, ELLs are able to develop English proficiency and access the content curriculum. However, the biggest challenge lies in the "moving target" nature of reaching grade level proficiency. As ELLs struggle to learn English in addition to content knowledge, their NES peers are progressing more rapidly, as they only have to focus on learning content knowledge. Therefore, in order for ELLs to catch up academically, they must simultaneously learn more content and language within one school year compared to their NES peers (Guerrero, 2004).

Another contributing factor to the achievement gap may be the variation in educational policies on the structure of language-programming and instructional delivery for ELLs. The allowance by the U.S. Department of Education for each state to formulate their own policies surrounding the identification, education, and assessment of ELLs results in a lack of consistency on educational policies. Additionally, NCLB (2002) has made it more difficult to implement bilingual programming, which gives rise to further variability in educational programming decisions. For example, while some states, such as Texas, continue to offer bilingual education options, other states, such as California, have passed legislation (Proposition 227) to allow only one year of intensive English-only education (California Department of Education, 1999) for ELLs. Both states have similar demographics, and represent the top two states with the highest proportion of ELLs, yet their educational policies are on opposite ends of the continuum.

Under NCLB's (2002) accountability system, all students must meet cutoffs on English high-stakes testing in order for schools to receive federal funding without regulatory controls on how these funds are spent. This provides further incentive to move away from bilingual education and push ELLs towards mainstreaming as soon as possible, regardless of student readiness. Even in sheltered English classrooms such as ESL, teachers may feel undue pressure to produce high standardized-test scores, which may translate into more emphasis on content mastery, without enough attention paid to promoting English acquisition. Despite accountability standards, in the 2007-2008 school year, only 11 states met self-set accountability goals for ELLs under NCLB (2002) in objectives for progress in English, and proficiency on state tests in reading and math

(Zehr, 2010). Since no comprehensive accountability systems exist within states, much less among states, there is not a clear way to measure overall ELL achievement under NCLB (2002). In order to appreciate the rationale and effects of current education policy, one must first understand the theoretical underpinnings of second language acquisition and instruction.

Perspectives of Language

Effective educational programs for language instruction are based on language-planning models that either address only language, or acknowledge the intertwined nature of language and culture. These models have defined objectives for the program, and stated language and cultural goals for what the learner should achieve. In some programs, this may be English acquisition and acculturation into mainstream culture, whereas other programs may strive for student bilingualism and biculturalism. Ruiz (1984) identified three perspectives on language: language as a problem, language as a right, and language as a resource. The language and cultural paradigms behind the organization of instructional programming for ELLs are based on these perspectives of language planning. Theories on how ELLs acquire a second language also contribute to instructional design decisions. Cummins (1979, 1981) delineates the difference between academic language and interpersonal language. Collier (1987) posits that 4-8 years are required for ELLs to acquire academic English. Through examining what is necessary for ELLs to develop fluency in a second language and the organizational rationale behind classroom models, one can better evaluate the effectiveness of existing instructional programming.

Language as a problem. The language as a problem perspective views limited English proficiency as a deficiency that must be addressed within the classroom (Ruiz, 1984). Historically, this is the predominant framework utilized in English-only educational programming for ELLs, and is the rationale behind compensatory language instruction program models. The learner's native language and culture are not utilized or even acknowledged through instruction, and are instead considered barriers to overcome. The ultimate goals of acquiring English-language proficiency and mainstream acculturation are emphasized through remedial English instruction. Educational policy decisions in Arizona, California, and Massachusetts, which only allow one year of intensive English instruction before mainstreaming ELLs in English-only monolingual classes, adhere to this perspective. Programming in Texas is less restrictive, but may still have elements that follow this perspective. Any instructional model that uses English as the only language of delivery also ascribes to this framework, examples include English as a Second Language and mainstream English-only classrooms.

Language as a right. This perspective views students' native language as a stepping-stone toward learning English (Ruiz, 1984). Instructional emphasis is placed on providing equal access to the curriculum through content and literacy instruction in students' native language until their English proficiency is adequately developed. Program models that adhere to this framework capitalize on students' native language proficiency as a way for ELLs to access the grade-level content curriculum while simultaneously acquiring academic English proficiency. The ultimate goals of this theoretical framework remain English competence and successful acculturation into

mainstream American society. In Texas, the program model that most closely adheres to this framework is transitional bilingual education. This model respects the linguistic rights of ELLs' native tongue, while moving towards the same language and cultural goals as the language as a problem framework.

Language as a resource. This last framework views students' native language and culture as an asset within the classroom, and it is the perspective that most appreciates the intertwined nature of language and culture (Ruiz, 1984). Program models that adopt this framework respect the linguistic human rights of students' native language, and seek to provide ELLs with an equitable education emphasizing bilingualism and biculturalism. These program goals focus on ***equal*** mastery of English as well as the native language. In Texas, dual-language immersion bilingual models follow this framework. The student makeup of these classrooms tends to be half ELLs and half NES peers for two-way dual language programs, or ELLs only in one-way dual language classrooms. The language of instruction generally follows a 50-50 distribution, with equal instruction in English and the native language.

Cognitive academic language proficiency. Cummins (1979, 1981) proposed a conceptualization of language as two distinct, independent continua known as Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency (CALP). BICS refer to the conversational interpersonal language that students use in everyday social interactions. CALP focuses on proficiency in academic language such as vocabulary used in classrooms, textbooks, and standardized tests. Cummins suggests that BICS tend to be developed through context-embedded

communication, which relies on communicative supports such as physical gestures and vocal inflections to help the listener make sense of the content. On the other hand, CALP is developed through context-reduced communication that does not offer clues to support understanding, such as textbooks. Therefore, the cognitive demand required for developing BICS is significantly lower than what is required for CALP.

The distinction between BICS and CALP is an important consideration in educational decision-making for ELLs. Although ELLs are able to develop peer-appropriate BICS in as little as six months to two years, it takes between 4-8 years on average to develop CALP (Collier, 1987). Hakuta, Butler, and Witt (2000) supported the finding that while academic English proficiency can take 4-7 years to develop; English oral proficiency only takes up to 3-5 years to develop. Students with developed BICS but emergent CALP may be erroneously mainstreamed from ESL or bilingual education into monolingual English classrooms before they are ready. Furthermore, with policy decisions pressuring schools to transition to English-only instruction as early as possible, more ELLs run the risk of being mainstreamed before the necessary timeframe for CALP development. More sensible policy will consider students' full array of academic needs and set aside the entire elementary school years as a realistic range in which English acquisition may be accomplished (Hakuta, Butler & Witt, 2000).

Language interdependence hypothesis. Collier (1987) demonstrated that ELLs between the ages of eight and 11 are able to gain CALP in the shortest amount of time compared to ELLs from different age ranges. These students were able to achieve the 50th percentile on national norms across content areas in only 2-5 years, whereas younger

ELLs took an additional 1-3 years to reach proficiency, and older ELLs required up to 6-8 years to reach grade-level norms. Cummins' language interdependence hypothesis (1979) suggests a common underlying proficiency is responsible for the development of CALP in different languages. As students develop CALP in their primary language (L1), the transfer of these cognitive skills to a second language (L2) occurs more readily. However, it should be noted this transfer is only relevant when the second language is structurally similar to the first, such as the case between Spanish and English. This may account for why late-elementary students are able to reach grade-level proficiency at a quicker rate than younger learners.

Cummins' Dual Iceberg Model (1981) for BICS and CALP (Figure 2.1) visually illustrates the level of understanding for both languages as students move from surface to deeper understanding. The common underlying proficiency serves as the central operating system responsible for the cognitively demanding tasks of CALP such as abstract reasoning in both languages. It is also responsible for making connections between first and second languages by utilizing existing linguistic knowledge of the student's similar native language.

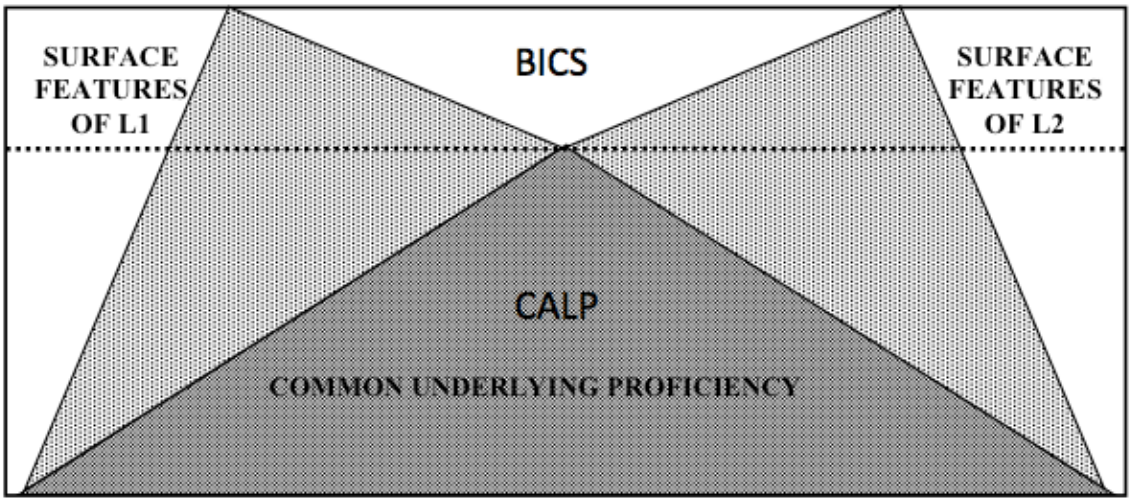

Figure 1. Cummins "dual-iceberg" representation of bilingual proficiency (1981).

Adapted from: Cummins, J. (1981). The role of primary language development in promoting success for language minority students. In California Department of Education (Ed.), Schooling and Language Minority Students: A Theoretical Framework, 24. Los Angeles, CA.

There may be important educational policy implications following the language interdependence hypothesis. This theory suggests that ELLs should remain in bilingual education long enough to develop CALP in their native language for the optimal transfer of language skills to the second language. This implication suggests that certain program models with an emphasis on native language proficiency and bilingualism may be more effective than English-only programs from a theoretical perspective.

Instructional Program Models for ELLs

The Bilingual Education Act of 1968 mandated that differentiated and appropriate educational opportunities be available for ELLs. As a result, diverse program models emerged to meet the educational and linguistic needs of these students (San Miguel, 2004; Valencia, 2008). Genesee (1999) proposes six broad categories of educational programming that serve as the basis for instruction of ELLs: sheltered instruction in

English, newcomer programs, transitional bilingual, developmental bilingual, second language immersion, and two-way immersion. The availability of these program models vary by state and student population, yet all instructional models for ELLs in the U.S. fit into one of these categories. Educational goals and language of instruction vary by program, as do student characteristics. Table 1 outlines Genesee's categories of educational programming. The programs shaded in this table are prominent in Texas and explored in depth. Each program is aligned with one of Ruiz's (1984) three perspectives on language planning, with differing degrees of emphasis on bilingualism and biculturalism. Not all of these program models are theoretically appropriate for ELLs, however all have been used in practice to some extent. Since the focus of this research is on Texas education policies, only the relevant programs models available in Texas will be reviewed and included in this study's methodology. This section describes the characteristics of available instructional programming, while the educational implications on achievement and student outcomes will be further discussed later.

Table 1

Descriptive Summary of Language Instructional Program Models

	Program Goals	Student Characteristics	Grades Served	Length of Student Participation	Language of Instruction
ENGLISH AS A SECOND LANGUAGE	Academic English proficiency; integration into mainstream American culture	Limited or no English; mix of language/cultural backgrounds; may have mix of ELL and NES students	All grades	Typically 1-3 years, longer as needed	English
NEWCOMER PROGRAMS	English proficiency; integration into mainstream American culture	Limited or no English; low literacy; recent immigrants; mix of language/cultural backgrounds	K-12, most prevalent in 6-12	Typically 1-3 semesters	English or native language
TRANSITIONAL BILINGUAL	Transition to all-English instruction; integration into mainstream American culture	Limited or no English; all students have same native language	Elementary grades with entry in K-2	2-4 years	English and native language, more English as students progress
DEVELOPMENTAL BILINGUAL	Bilingualism; biculturalism	Limited or no English; all students have same native language	Elementary grades with entry in K-2	Usually 6 years, preferably 12 years	English and native language
FOREIGN/SECOND LANGUAGE IMMERSION	Bilingualism; appreciation of new culture while maintaining mainstream culture	Native English speakers	K-8, preferably K-12; entry in K-1	Usually 6 years, preferably 12 years	Second language with English adaptations
TWO-WAY IMMERSION	Bilingualism; biculturalism	Mix of NES and ELL students	K-8, preferably K-12; entry in K-1	Usually 6 years, preferably 12 years	English and minority language

Adapted from: Genesee, F. (Ed.) (1999). Program alternatives for linguistically diverse students. Educational Practice Report 1. 3. Berkeley, CA: Center for Research on Education, Diversity & Excellence.

English-language monolingual program. This program provides English-only instruction without special language support (Genesee, 1999), and follows the "language as a problem" perspective of language education (Ruiz, 1984). Before bilingual and ESL options were available, monolingual English was the default program model for all students. This is not technically considered a program model for ELLs, as there are no specific language accommodations used in these classrooms. However, parents of eligible ELLs have the option of refusing bilingual and ESL services when enrolling their children in school, thereby placing them in these mainstream English settings. ELLs in this environment are educated alongside NES peers by a teacher who does not hold additional certification beyond the content area. Thomas and Collier (2002) identified ELLs enrolled in this setting as having the poorest outcomes in reading and math achievement by 5th grade. The largest number of dropouts also came from this group of students (Thomas & Collier, 2002). In the current study, achievement outcomes of ELLs who did not receive services will be examined and demographically similar non-ELL students enrolled in mainstream English classrooms will serve as the comparison group for student achievement outcomes.

English as a second language. This type of programming provides English-only instruction with scaffolded language support, and is generally taught by English as a Second Language (ESL) certified teachers (Genesee, 1999). The key characteristic is that all instruction is delivered in English only. Thus, the primary language goal is for ELLs to develop academic English proficiency. ELLs enrolled in this program model may receive instruction through a variety of venues. One method is the pullout model, where

the student is educated amongst NES peers using no special accommodated language instruction for the majority of the school day. These ELLs also attend a specialized language instruction classroom taught by a certified ESL teacher for designated time periods built into their schedule. Alternately, ESL programs may follow an inclusion model, where students remain in one classroom taught by an ESL teacher, and content instruction is scaffolded so that ELLs learn both language and content knowledge simultaneously. The student population in these classrooms may be ELLs only or a mix between ELL and NES students. Like monolingual English-only instruction, ESL program models adhere to the "language as a problem" theoretical framework (Ruiz, 1984), where students' lack of English proficiency is viewed as a deficiency, and native language proficiency is largely ignored.

Bilingual education programs. Bilingual education in the U.S. has evolved into three separate program models: transitional, developmental, and two-way immersion (Genesee, 1999). All three instructional models utilize both English and students' native-language (most commonly Spanish) to deliver content. The ratio of English to Spanish varies by program, but all provide instruction in Spanish at least 50% of the school day. Currently in Texas, there exists early and late-exit transitional bilingual education, where the classroom is composed of all Spanish-speaking ELLs, as well as two-way immersion, which are ideally composed of equal numbers of NES and ELL students.

Transitional bilingual education program. As its name suggests, transitional bilingual programs have the ultimate goal of transitioning ELLs to mainstream English-only instruction (Genesee, 1999). These programs focus on student integration into

American culture, and development of academic English proficiency. The teacher's use of Spanish in instruction is perceived as a stepping-stone to the ultimate goal. This program model ascribes to the "language as a right" theoretical framework (Ruiz, 1984), where students' native language proficiency is capitalized on as a way to develop English proficiency. The classroom makeup is homogenous ELLs beginning with little to no English proficiency. These programs may start out in early childhood classrooms as 90-10 models, with 90% of instruction in Spanish, and 10% English supplement. However, by the time students reach first or second grade, instruction is gradually shifted to a 50-50 model, with equal time spent on English and Spanish.

In Texas, early-exit transitional programs have the goal of mainstreaming students to English-only instruction in 2-5 years, while late-exit programs aim to transition these students in 6-7 years. Prior to 2010, Texas also offered developmental bilingual programs, with a greater emphasis on bilingualism than transitional programs. While the duration of programming for late-exit transitional and developmental bilingual are equal, it appears the policy emphasis in Texas has shifted away from bilingualism and towards accessing the English-only curriculum through abolishing the developmental bilingual model. Students typically enroll in transitional programs between pre-kindergarten and second grade, and are exited from bilingual education by the end of elementary school. The average length of program enrollment is 2-4 years. Under the policy changes of Texas House Bill 3 (HB3) (2009), 5th grade is the last year a Spanish version of the statewide high-stakes assessment is available. The statewide assessment was previously the Texas Assessment of Knowledge and Skills (TAKS), but was

replaced by the more rigorous State of Texas Assessments of Academic Readiness (STAAR) in Spring 2012. All middle and high school students are assessed in English regardless of a student's English language proficiency. As educational policy continue to shift in Texas, more bilingual education programs are following the early-exit transitional model, with students exiting bilingual education several years before middle school.

Two-way immersion bilingual education program. Two-way immersion (also known as dual-language bilingual) programs are a relatively new instructional model, and may offer the solution to the problems presented by both English-only instruction and traditional bilingual education. Dual-language programs aim for bilingualism and biculturalism for ELLs (Genesee, 1999). This model follows the "language as a resource" theoretical framework (Ruiz, 1984), which values students' native language, and seeks to continuously develop native language skills in addition to English proficiency. Additionally, the classroom makeup consists of equal numbers of ELLs and NES students, and instruction is also aimed at NES students developing Spanish-language proficiency. The classroom culture is structured to promote interaction and mutual respect between both groups of learners. In theory, students enroll between kindergarten and 1st grade and remain in the program through 8th grade, ideally through high school. The longer duration of this program yields benefits for all learners (Genesee, 1999; Thomas & Collier, 2002). ELLs are able to develop English proficiency at a more relaxed and natural pace, while forming genuine friendships with peers from the mainstream culture. For NES students, this program provides the opportunity to develop bilingualism without leaving the country, and also affords them the chance to participate in and

develop a rich understanding of a minority culture. Several research studies support this program model as producing the highest long-term student academic achievement in English of all the instructional models for ELLs (Thomas & Collier, 2002; Krashen, 2004). NES students in this program also perform comparably to their mainstreamed English-only peers.

In Texas, the duration of dual-language immersion programs is between 6-7 years, the same as late-exit transitional bilingual programs, meaning that all bilingual education supports end by the time Texas ELLs begin middle school and are expected to take high-stakes assessments in English. It is unknown whether this shorter length of dual-language immersion programming yields the same high achievement results as dual-language immersion that extends through secondary education.

Effects of Instructional Models on Academic Achievement

Under optimal academic circumstances, ELLs are able to perform at a rate commensurate with their NES peers. This topic has garnered much research attention in the past few decades, and patterns of achievement are beginning to emerge. There are many factors that predict future achievement in ELLs, such as instructional model, classroom characteristics, and length of schooling. The current study includes some of these factors as additional independent variables to test their significance as predictors of future academic achievement.

Thomas and Collier (2002) conducted a national longitudinal study comparing the long-term academic achievement of ELLs enrolled in diverse classroom program models. Two-way dual language bilingual education programs emerged as the most promising

model, for both ELLs and their NES classmates. ELLs in these programs performed significantly higher in 5th grade on all subject areas compared to ELLs enrolled in ESL, developmental bilingual and transitional bilingual programs. Krashen (2004) reviewed the existing literature on ELLs' academic English attainment after enrollment in two-way bilingual programs. His results indicate that this program model produces generally positive but variable attainment of academic English.

Although dual-language programs appeared the most favorable, both transitional and developmental bilingual education produced ELLs with higher academic achievement than those enrolled in ESL models (Mora, Wink, & Wink, 2001; Thomas & Collier, 1997; 2002), with developmental bilingual programs showing superiority over transitional bilingual education (Rolstad, Mahoney, & Glass, 2005). Goldenberg (2008) compared two large meta-analytic literature reviews on educating ELLs conducted by the National Literacy Panel on Language-Minority Children and Youth (NLP) and Center for Research Design on Education, Diversity and Excellence (CREDE). Both meta-analyses concluded that primary language instruction either prior to or simultaneously as learning a second language leads to higher levels of ELL achievement in English (August & Shanahan, 2006; Genesee, Lindholm-Leary, Saunders, & Christian, 2006), lending further support for the benefits of bilingual education.

Other research demonstrates less clear advantages between programs, especially across a broad range of skills. Slavin, Madden, Calderon, Chamberlain, and Hennessy (2011) compared multiyear reading and language outcomes of Spanish-speaking ELLs randomly assigned to transitional bilingual or structured English immersion programs. In

4th grade, after bilingual students were transitioned to English-only instruction, no differences were found in English reading outcomes. Other studies comparing reading achievement in 3rd grade after three years of enrollment in different instructional programming found little difference in student reading skills (Galvan-Luis, 2010; Nakamoto, Lindsey, & Manis, 2012). However, students enrolled in bilingual education performed better on measures of oral English language proficiency, with students in dual-language bilingual programs showing the most improvement (Galvan-Luis, 2010). Taken together, the literature suggests that bilingual programming shows promising results over English-only instruction, but research still lacks conclusive evidence that there is one best program model for ELLs.

The amount of language programming ELLs receive also contributes to their future achievement (Collier, 1987; de Jong 2004; Dixon, 2012; Goldenberg 2008; Thomas & Collier, 2002). Thomas and Collier (2002) found that students who receive at least 5-6 years of dual language schooling in the U.S. are able to perform on grade-level in English by middle school. Dixon (2012) concluded that ELLs without disabilities take a mean 7.354 years to reach English proficiency, while the CREDE and NLP studies found that it takes at least six years for most students to progress from beginning to native-like proficiency in a second language (Goldenberg, 2008). These results are consistent with Collier's (1987) assertion that ELLs require between 4-8 years on average to develop academic English proficiency. Cummins' (1979, 1981) language interdependence hypothesis may also account for this finding, as students in dual language programming are developing academic proficiency in both their native language

and English. The common underlying proficiency that supports native language development also develops academic English (August & Hakuta, 1997; Cummins 1979; Cummins 1981).

In addition to amount of language instruction received in the U.S., the amount of prior schooling also contributes to higher achievement (Orozco, 2012; Thomas & Collier, 2002). Immigrant ELLs who received continuous, strong, grade-level schooling prior to enrolling in the U.S. had higher academic achievement in English. Furthermore, the detrimental effects of low SES had less influence on ELLs who received four or more years of prior schooling; the benefits of prior schooling were able to offset these negative effects (Thomas & Collier, 2002). Latino ELLs enrolled in pre-kindergarten programming in either English or Spanish had higher achievement in 3rd grade compared to students who did not receive early childhood education (Orozco, 2012). These results suggest that earlier age of school enrollment, and subsequently more years of education in either a student's primary language or English, are predictive of higher achievement outcomes.

There are additional characteristics of the classroom environment that may promote high academic achievement in ELLs. When the learning environment is culturally sensitive, supportive, and integrates students' home and school identities, ELLs feel valued and are able to achieve more academically (Curtin, 2005; Garcia, 1991; Iddings & Katz, 2007). High quality instruction that incorporates language and literacy instruction, cooperative learning, professional development for all staff, progress monitoring, and organized school leadership are also predictors of ELL success

(Calderon, Slavin & Sanchez, 2011). Positive parental relationships with schools similarly promote learning, as students may feel more invested in their education (Garcia, 1991). Finally, when comprehensive language instruction is provided in a context-embedded manner, such as through scaffolded grade-level content instruction, ELLs are better able to learn both content knowledge and the English language (Echevarria, Short, & Powers, 2006). These models have been shown to be more effective than de-contextualized language instruction that is separate from content instruction, such as pull-out ESL program models (Callahan, 2006).

Long-Term ELLs

Despite the availability of language programming and the empirical evidence supporting its effectiveness, there remains a large group of ELLs who perennially struggle to attain English language proficiency. Long-term ELL (LTELL) is a subset of the ELL population who require five or more years to reach English proficiency (Menken & Kleyn, 2009; Olsen, 2010). These learners are typically adolescents, found in grades 6-12, with a high risk of dropout, grade retention, and poor postsecondary outcomes (Menken & Kleyn, 2009; Olsen, 2010). This unique group is often overlooked in the classroom and understudied within the ELL literature, yet they constitute 59% of secondary ELLs in the country, and one-third of all ELLs in California (Olsen, 2010). A deeper understanding of the learner characteristics and achievement outcomes of this population may facilitate a better understanding of why some ELLs are able to attain English proficiency, while others remain stuck in ELL categorization year after year.

LTELLs tend to be orally bilingual and sound like native English speakers. However, they lack literacy skills in their native language, and their academic English literacy skills (CALP) are not as well developed as their oral language (BICS) (Menken & Kleyn, 2009). Many LTELLs are U.S. born, having spent nine or more years in U.S. schools without developing sufficient academic English. In fact, Foreign-born ELLs tend to develop academic English faster than U.S. born LTELLs (Slama, 2012). Menken and Kleyn (2009) point out that the U.S.-born label can be misleading, as many LTELLs are transnational students who have moved back and forth between the U.S. and the family country of origin, attending school in both countries. These transnational students constitute one of the two main subsets of LTELLs. The other subset are learners who have received inconsistent schooling in the U.S., frequently changing schools and moving between language programming models and mainstream English classrooms, possibly due to a migrant lifestyle or poverty (Menken & Kleyn, 2009; Calderon, Slavin & Sanchez, 2011).

Olsen (2010) outlines some of the factors contributing to the high number of LTELLs. These students often do not receive any language-instructional programming, or are enrolled in weak or poorly implemented programs, yielding limited access to the full curriculum. LTELLs tend to develop habits of non-engagement, learned passivity and invisibility within the school context, resulting in self-selected social segregation, and linguistic isolation (Olsen, 2010). The academic outcomes for LTELLs are particularly grim. They experience significant deficits in reading and writing, have gaps in their background academic knowledge, and are stuck at intermediate levels of English

language proficiency or below (Olsen, 2010). Additionally, LTELLs struggle in all content areas requiring literacy, receive poor grades and experience frequent grade retention, which are all factors that place them at greater risk for dropout (Menken & Kleyn, 2009).

LTELLs have different linguistic and academic needs than those of newly arrived ELLs, yet language programming at the secondary level is typically intended for new arrivals (Menken & Kleyn, 2009). Thus, these students are often inappropriately placed in newcomer programs or mainstream education, and taught by underprepared teachers (Olson, 2009). Other barriers to achievement include limited access to the full curriculum due to low literacy skills and a lack of knowledge by students and their families that they are underachieving (Olson, 2009). Recommendations for LTELLs focus on school leadership, progress monitoring, literacy development and instructional techniques within the secondary classroom (Calderon, Slavin & Sanchez, 2011). It is clear that adolescent ELLs who have received inconsistent schooling are at higher risk for educational failure, but less is understood about the characteristics of LTELLs afforded similar opportunities as former-ELLs. Why are some ELLs able to reach English proficiency by 6th grade, while others continue to struggle? This study aims to fill that research gap by exploring the varying factors between LTELLs and former-ELLs, the latter of which have much more promising achievement outcomes.

Former-ELLs

The achievement gap between current ELLs and their NES peers is already well established (U.S. Department of Education, 2009a; 2009b; 2010a; 2010b). However,

research on what happens to this achievement gap ***after*** these ELLs complete language programming and are mainstreamed into English-only instruction is only beginning to emerge. Much of the research examines the academic trajectories of former-ELLs compared to their NES and LTELL peers without taking into consideration the classroom models of instruction. The comparison of achievement outcomes for former-ELLs enrolled in differential program models may have strong implications for the program evaluation of real-world language instructional models.

A growing body of recent research supports positive long-term achievement outcomes for former-ELLs (Ardasheva, Tretter, & Kinny, 2012; Flores, Batalova, & Fix, 2012; Halle et al., 2012; Kim, Herman, & National Center for Research on Evaluation, 2009; Kim, Herman, & National Center for Research on Evaluation, 2012). Flores, Batalova and Fix (2012) conducted a large-scale longitudinal study following several groups of demographically diverse ELLs in Texas from 1st grade (1995) through 12th grade (2006). Results demonstrated that the cohort of former-ELLs who progressed through school without grade retention or dropping out had higher achievement (TAKS scores, graduation rates, postsecondary educational attainment) than all other ELLs, although still lagging behind their NES peers on all measures of achievement. However, former-ELLs who exited from language programming after three years had the best results of meeting reading and math proficiency standards, surpassing the performance of other ELLs and even their NES peers. Other studies also supported that former-ELLs assessed two or more years after reclassification tend to outperform current ELL and NES peers in both math and reading (Ardasheva, Tretter, & Kinny, 2012; Kim, Herman,

& National Center for Research on Evaluation, 2009), although this was not a consistent finding across the literature.

The age and timing of reclassification appears to be another variable impacting outcome measures. Former-ELLs reclassified at an earlier age, in kindergarten, kept pace with their NES peers on academic measures, and tended to have better social and behavioral outcomes when compared to NES peers and students who were reclassified at later ages (Halle et al., 2012). Kim, Herman and National Center for Research on Evaluation (2012) discovered that students reclassified around the end of elementary school, in 4th through 6th grades, tended to continue to catch up with their non-ELL peers, but not at a level sufficient to gain equal footing. Additionally, gaps between former-ELL and non-ELL students increased significantly with each grade of reclassification, further supporting that students ready to be mainstreamed earlier have better achievement outcomes.

Prior studies have attempted to compare former-ELLs educated in different classroom models of instruction (de Jong, 2004, Thomas & Collier, 2002) with varying degrees of success. De Jong's (2004) research compared achievement test scores of students enrolled in ESL and bilingual education and their NES peers in 4th and 8th grade. Results suggest that students enrolled in ESL have higher long-term achievement than students enrolled in bilingual education. However, the author failed to control for demographic variables that may impact achievement such as ethnicity, age of enrollment, and SES. Thomas and Collier's nationwide sample of language program outcomes provides a more comprehensive study of long-term achievement patterns while

controlling for demographic variables across different school districts. Their results support bilingual education programming as having the most promising outcomes of all the language program models studied. Nevertheless, this study was conducted before NCLB (2002) dramatically changed the instructional and assessment landscape.

The current study aims to add to this growing body of literature by comparing differential student achievement outcomes after NCLB (2002) of students enrolled in different instructional models, while controlling for demographic variables, with a focus on low-SES, Latino learners. The literature suggests that while the majority of Latino ELLs come from low-income backgrounds, there is greater heterogeneity among these students compared to other groups of ELLs (Flores, Batalova, & Fix, 2012; Kim, Herman, & National Center for Research on Evaluation, 2009). Whereas some are able to catch up with their non-ELL peers after reclassification, others become long-term ELLs unable to attain English proficiency. This variability makes this diverse group of learners an especially relevant and interesting population of study. Figure 2 visually illustrates the various ELL categorizations as they advance through the U.S. school system.

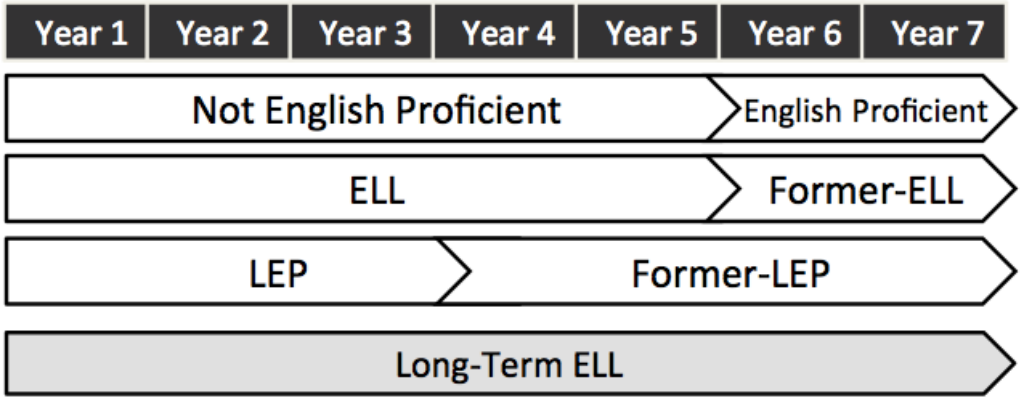

Figure 2. Categorization of ELL identification in Texas after years of U.S. schooling*. **Note: This visual provides a representation of the average length of U.S. schooling received before ELLs are reclassified as English proficient. There exists great variability between learners on actual time required to achieve English proficiency.*

Focus on Texas

Texas is a state with one of the longest running high-stakes testing policy in the U.S. Beginning in 1984, Texas legislature first mandated making graduation contingent on students passing exams in math and reading. In the 1990's, then governor George W. Bush established the first set of high-stakes assessments (the Texas Assessment of Academic Skills), which held school districts accountable for student test scores, including cash awards for high performance and sanctions for low ratings. When Bush became president, he signed the NCLB (2002) act into law, which is based on his previous model of high-stakes testing in Texas and mandates all states conduct high-stakes standardized assessments (Hursh, 2005; Vasquez-Heilig & Darling-Hammond, 2008). As historically illustrated, Texas educational policies have significant implications for policy decisions on a national level. Furthermore, the population of ELLs in Texas is the second largest in the country, surpassed only by California. Therefore, these students represent a significant proportion of ELLs nationwide. Their educational outcomes will

be of interest to other states that look towards Texas as an exemplar for the identification, education, and assessment of ELLs.

Another reason for focusing on Texas is the abundant availability of student-level data. Texas is the national leader in creating longitudinal data with a system of individual student identifiers, which allows for the tracking of individual student's progress and other achievement outcomes (Losen, Orfield, & Balfanz, 2006). These individual student identifiers also permit the tracking of students across public schools and school districts within Texas. This statewide database is known as the Public Education Information Management System (PEIMS) and will be further discussed in the methods section. The existence of this data system affords researchers with the capacity for accurately analyzing and reporting a host of achievement outcomes that would not be possible in other states.

Statement of Problem and Purpose

ELLs are among the fastest growing student populations in the U.S. public school system, with Texas ranking second highest of all states for the number of LEP students (U.S. Department of Education, 2006). The need to identify, educate and assess ELL students presents a formidable challenge for policy makers and educators alike, particularly given the requirements of NCLB (2002), which mandates that all students participate in high-stakes testing. The U.S. Department of Education allows each state to devise their own rules and regulations surrounding the identification and assessment of ELLs. This provision results in inconsistent laws across states on the types of assessments used, models of English language instruction, and allowable accommodations.

In 2012, ELLs represented 17% of the total Texas student population, which translates to over 838,000 ELLs spanning pre-kindergarten through 12th grade (Texas Education Agency, 2012). As a state with one of the highest numbers of ELLs, Texas bears strong responsibility to establish effective policy on the identification, education, and assessment of these students. It is also a state with the longest running history of high-stakes testing, with a wealth of student-level data. As ELL populations continue to grow within public school systems, other states will look to Texas as an exemplar for ELL education models. Therefore, it is important to determine the effectiveness of current policy practices as measured by student achievement outcomes, and identify the factors which contribute to future student success beginning in elementary school.

The primary interest of this study is in examining the long-term academic achievement of two groups of demographically similar ELLs across instructional program models, ESL and bilingual education, and in comparison to their non-ELL peers. Achievement data from a cohort of participants across the 10 major urban school districts in Texas who began school before 3rd grade were retrospectively examined at two points in their public education: 7th, and 9th grade. The study followed students enrolled in 3rd grade during the 2003-2004 school year until 9th grade during the 2009-2010 school year. To examine the impacts of instructional model on future achievement, ELL participants were categorized into three groups based on primary classroom education model received: ESL, bilingual education, and ESL/bilingual for students who attended equal number of years in both programs. A third group of ELL students who did not receive any classroom model services (i.e., participated in mainstream English-only education)

during the study years emerged during the analysis process and their achievement results were compared as well. Achievement results from a comparison group of demographically similar non-ELL students were also examined.

From this existing sample, ELL participants were re-categorized into two groups to address differential achievement between types of ELLs and their non-ELL peers. The groups are: former-ELLs who started school as ELL but were reclassified as English proficient before the start of secondary education (6th grade), and long-term ELLs who also started school as ELL but were either reclassified as English proficient after 5th grade, or remained ELL through the end of the study (9th grade). Individual characteristics from the two types of ELLs were examined in an exploratory analysis to compare group differences on independent variables (i.e., immigrant status, primary classroom education model received), in comparison to their non-ELL peers, and controlling for English proficiency level in 3rd grade. These variables are demonstrated to be significant within the literature as factors affecting achievement. District of student enrollment was included in all analyses to examine the uniformity of educational practices across the state.

Research Questions

The following research questions were formulated:

Research question 1. Controlling for demographics (e.g., Latino, low-SES), and compared with non-ELL peers, are group mean differences present between students enrolled in various primary language program models (ESL vs. bilingual vs. equal years

ESL/bilingual vs. no services) during elementary years on ELL student academic achievement in secondary grades?

Research question 2. How do former-ELLs (defined as reclassified to English proficient in 5th grade or earlier) perform on measures of academic achievement in 7th and 9th grade compared to long-term ELLs (defined as reclassified to English proficient after 5th grade, or not reclassified by end of study) and their non-ELL counterparts?

Research question 3. Assuming that hypothesis 2 is confirmed and former-ELLs have higher achievement outcomes than long-term ELLs in 7th and 9th grade, what individual differences between groups significantly contribute to differential achievement outcomes?

Chapter Three: Method

Changes from Proposed Study

The study originally proposed to examine a student sample from one urban school district in Texas. Given that the Texas Education Research Center (Texas ERC) database contains a wealth of student-level data spanning Texas and at the suggestion of the Texas ERC Advisory Board, I decided to expand the study's sample to examine data from all 10 major urban school districts in Texas during the study's timeframe. This decision was made in hopes of increasing statistical power and producing more thorough research, with a larger sample size, that may be better generalized to students across the state. A secondary purpose was to examine the educational uniformity of district practices relating to the education of ELLs across demographically similar school districts in Texas.

This study originally proposed to examine student achievement at four points during their schooling rather than two, as the proposal utilized a growth curve analysis within a hierarchical linear framework. Based on changes to the study's design, the data structure, and available resources, a different data analysis was selected, which did not require as many repeated measures. As the primary interest of the study was to examine long-term student achievement, 7th and 9th grade achievement scores remained the focus of the analyses. The original study proposed to examine two types of bilingual education: early-exit and late-exit. The available data did not delineate the types of bilingual education at the beginning of the study, therefore bilingual education was examined as a whole.

The original study proposed to utilize a comparison group of demographically similar NES peers. However, data available to the researcher began at this cohort of students' 2nd grade year (2002-2003), and it was not possible to deduce students' ELL status prior to this year. Thus, the comparison group consisted of students who were not ELL as of 2nd grade, and are referred to as "non-ELL" rather than NES students. In order to capture ELLs as thoroughly as possible, students identified as ELL in 2nd grade remained in this study as part of the ELL sample; however, some of these students had already exited language programming by the start of the study's timeframe in 3rd grade. Regarding the third research question's exploratory analyses, the independent variables examined were limited by constraints in data availability and methodology. Several variables were also confounded with the independent variable of ELL type, which takes into account duration of ELL status. As a result, duration of program enrollment, age of reclassification, and grade of exit from language instruction, originally proposed as independent variables of interest, were not ultimately included in this current study.

During the analysis process, two new groups of classroom language models emerged from the data that I did not originally account for: ELL students who received equal years of bilingual education and ESL, and ELLs who did not receive any language programming during the study's timeframe (2003-2010). These new groups are incorporated below in Research Question 1 and the accompanying hypotheses and rationale.

Participants

Data set. This study involved a secondary data analysis of individual student-level data derived from Texas Education Agency's Public Education Information Management System (PEIMS). Student data between 2003-2010 were examined from the 10 major urban school districts in Texas as of the 2009-2010 school year, the last year of the current study. The Texas Education Agency (2011) classifies a district as "major urban" if it meets the following criteria: (a) located in a county with a population of at least 750,000; (b) its enrollment is the largest in the county or at least 75% of the largest district enrollment in the county; and (c) at least 35% of enrolled students are economically disadvantaged, measured by the student's eligibility for free or reduced-price meals. As of 2015, there are currently 11 districts in Texas that meet these standards. In 2009, there were 10 major urban school districts in Texas that are included in this study: Dallas, Houston, El Paso, Austin, Fort Worth, North East, Northside, San Antonio, Ysleta, and Arlington. For the purposes of this study, district names will not be identified in the results; rather they are arbitrarily numbered 1-10.

The PEIMS is a statewide database supervised by the Texas Education Agency (TEA), and encompasses all data requested and received by TEA about public education in Texas. The database includes individual student information on: enrollment, demographics, special program participation, attendance, discipline, and school leaver data such as withdrawn students, graduates, and dropouts. Additionally, the PEIMS is linked to a separate database containing student results on the Texas Assessment of Knowledge and Skills (TAKS), the statewide high-stakes assessment.

This study utilized data from the Texas Education Research Center (Texas ERC) at The University of Texas at Austin. This is a virtual research center that enables scientific inquiry through a clearinghouse of statewide information. It is an independent, non-partisan, and non-profit organization focused on data-informed decision-making for Texas education. The Texas ERC provides access to longitudinal data from TEA spanning pre-kindergarten through higher education. Contained in this repository database are PEIMS and TAKS data. Specifically, this database contains student-level data on enrollment, attendance, course completion, graduation, leaver/dropout, and discipline, ranging from 1993-2010. Regarding assessment results, the Texas ERC contains TAKS data for all grades and subjects in both English and Spanish spanning 2003-2010. Additionally, this database contains Texas English Language Proficiency Assessment Systems (TELPAS) results from 2004-2009. This clearinghouse of information served as the sole data source for the current study. Researchers who are interested in accessing this information must submit their request for data review in writing, and follow the agency's protocols on security and privacy. Permission to access this database was granted after a telephonic advisory board meeting conducted on May 21, 2014. The board granted access to specific data files outlined in the Texas ERC research proposal (Appendix A) and recommended that the study sample encompass student data from multiple school districts. Sampling procedures were thus modified to meet this requirement and are outlined below.

Sampling. The current research was a retrospective longitudinal study examining student data from a cohort of 18,188 students who met sampling criteria, enrolled in the

10 major urban Texas school districts from 3rd through 9th grade between the 2003-2004 and 2009-2010 school years. To address the issue of school mobility, students who moved between schools and major urban districts remained in the study, identified by their district of attendance in 3rd grade. Students who moved out of the target districts for any year of the study were excluded from the sample.

Students who met the following criteria were included in the study: Latino, enrolled in 3rd grade during the 2003-2004 school year, receiving free or reduced meals, identified as ELL in 2nd grade, attendance in one or more of the 10 major urban school districts for the duration of the study, and Spanish as the primary household language indicated by the Home Language Survey (HLS). Students in the comparison group met the same demographic criteria of race, grade, district of attendance, and free or reduced lunch. These students were not identified as ELL in 2nd grade, had primary household languages of Spanish and/or English, and received mainstream English-only education.

To address the first research question, all students in Texas who met the target criteria were included in the analysis. Variables were derived to categorize students into groups based on primary language programming received (e.g., ESL, bilingual, equal ESL/bilingual, no services, non-ELL). From this existing sample, ELL participants were re-categorized into two groups to address the remaining research questions: former-ELLs who started school as ELL but were reclassified as English proficient before the start of secondary education (6th grade), and long-term ELLs who also started school as ELL but were either reclassified as English proficient after 5th grade, or remained ELL for the duration of the study.

All students in this study were enrolled in general education throughout the study's duration and special education enrollment at any time during the study's timeframe was an exclusionary factor. However, descriptive data on special education students within the ELL population are reported and explored further in the discussion chapter, as there exists significant overlap between ELL and special education populations (Texas Education Agency, 2009; 2012a). Students who received special education in 2nd grade but exited by 3rd grade, prior to the start of the study, remained in the sample.

Variables and Instrumentation

The dependent variables for all research questions were measures of academic achievement, as represented by TAKS scores in reading and math. For each research question, TAKS scores were examined in 7th and 9th grades. Only scores from the first administration of the assessment, which encompasses the majority of test-taking students, were included in this study. There are several additional TAKS administration days that serve as make-up test dates for students who were absent and as retake opportunities for students who failed the original administration. These additional scores were not included in the analyses due to inconsistent reporting in the available data (i.e., delineating retake scores versus make-up scores). Student ELL status is determined on an individual basis by the Language Proficiency Assessment Committee (LPAC) on each campus (Texas Education Agency, 2012b). Students who qualify as ELL have a home language that is not English, and a beginning TELPAS score. The LPAC also determines when students are exited from ELL status, and become former-ELL students. This is achieved after

three years of schooling in the U.S. for immigrant ELLs and a TELPAS score of advanced. Due to limited data availability, English language proficiency (ELP) at baseline only will be included as a covariate variable, and measured through student TELPAS scores, a statewide assessment that is administered annually to ELLs. Research question 3 examines the possible interaction between language programming received and ELL-type while controlling for baseline ELP level. This research question also examines the effects of immigrant status, described within the literature as a relevant predictor of ELL achievement. Table 2 outlines the variables to be collected.

Table 2

Variables To Be Collected at Each Time Point

3rd Grade: 2004	**7th Grade: 2008**	**9th Grade: 2010**	**Additional Variables:**
TELPAS score	TAKS reading score	TAKS reading score	Immigrant Status – Yes/No
District of attendance	TAKS math score	TAKS math score	Primary classroom model received
			ELL type

Texas English Language Proficiency Assessment System (TELPAS).

Overview. The state of Texas uses the TELPAS as the statewide English Language Proficiency (ELP) assessment, and these scores at baseline were utilized in the current study as a covariate variable. This assessment plays a major role in educational decision-making, ranging from classification of ELLs, to curriculum planning and instruction (Abedi, 2007). In order to provide a fair and valid assessment for ELLs, and ensure an equitable educational opportunity for all students, the NCLB (2002) act

mandated reporting of Adequate Yearly Progress (AYP) for all students, including ELLs. This accountability requires the assessment of ELLs in two ways, reliable and valid measurement of ELP, as well as grade-level content knowledge.

The statewide ELP test of Texas is comprised of two components: the Reading Proficiency Test in English (RPTE) and the Texas Observation Protocol (TOP). TELPAS results have been used since 2005 to report the Annual Measurable Achievement Objective accountability measures as required by NCLB. The test measures ELLs' annual English progress in four domains: listening, speaking, reading, and writing. Results are used in combination with other measures (such as the TAKS) to inform instructional decisions for individual students.

Score Reporting. The TOP is scored holistically, with the teacher assigning a proficiency rating for the domains of listening, speaking, and writing. The RPTE is a multiple-choice, standards-based reading test. There are four TELPAS proficiency ratings: Beginning, Intermediate, Advanced, and Advanced High. Students are given a proficiency rating in each of the four language domains. To obtain a student's composite score, the proficiency rating is first converted to a number from 1 (Beginning) to 4 (Advanced High) for each language domain. Each domain rating is weighted, with reading given the most weight (75%) in the composite rating, followed by writing (15%), listening (5%) and speaking (5%) are given the least weight. The resulting weighted composite score ranges from 1 (ratings of Beginning in all language areas) to 4 (ratings of Advanced High in all language areas). In the current study, only the student's

composite score is used in the data analyses as a measure of the student's overall English Language Proficiency.

Psychometric Properties. Reliability and validity estimates for TELPAS are provided in the Student Assessment Division's Technical Digest (TEA, 2010). The Kuder-Richardson Formula 20 (KR20) was used to calculate the internal consistency reliability estimates. For the spring 2009 TELPAS reading tests, the internal consistency estimates were excellent, ranging from 0.93 to 0.96. This test also had a small standard error of measurement (SEM) of between two and three raw score points across grades. For the holistically rated domains of the TELPAS, evidence of high interrater reliability was collected through an audit process. Further analyses indicate the weighted TELPAS composite ratings have reliability estimates exceeding 0.89. The TELPAS is based on the state English Language Proficiency Standards (ELPS), which are part of the Texas Essential Knowledge and Skills (TEKS) curriculum. Both the TAKS and the TELPAS undergo a similar development process based on the TEKS curriculum. Texas conducted two studies in 2008-2009 to investigate student performance on TELPAS compared with student performance on the TAKS. Results indicate that as a student's TELPAS proficiency rating increases, so do their TAKS Reading and ELA scale scores. This strong relationship provides validity evidence supporting the TELPAS. This also suggests that TELPAS scores may be confounded with TAKS reading scores, which led to the decision to include the TELPAS score in this study as a covariate variable rather than an independent variable. Specific information on test reliability and validity can be found in the state's annual technical digest (TEA, 2010).

Texas Assessment of Knowledge and Skills (TAKS).

Overview. The TAKS was the high-stakes assessment of content knowledge for the state of Texas under NCLB (2002), between 2003 and 2011. Beginning in Spring 2012 the TAKS was replaced with the more rigorous State of Texas Assessments of Academic Readiness (STAAR) as the statewide high-stakes assessment. The TAKS was mandatory for the majority of students starting in 3rd grade. It is a criterion-referenced exam that is aligned with the TEKS curriculum, and measures students' ability to apply grade-level learned content knowledge. By law, all public school students are annually assessed with TAKS starting in 3rd grade across a variety of subjects including: math, reading, social studies, science, writing, and English language arts. Math and reading are assessed every year, whereas the other subjects are assessed less frequently on a predetermined schedule. In addition to the standard TAKS for each grade level, different versions of the TAKS are used to assess special student populations, including special education and LEP. The Spanish TAKS was developed to assess the content knowledge of ELLs enrolled in bilingual education. Until the 2009-2010 school year, Spanish TAKS was available for grades 3-6. As of the 2010-2011 school year, and as a result of Texas HB3 (2009), the Spanish TAKS is now only available for grades 3-5. For this study, English TAKS scores in math and reading were utilized as an indicator of academic achievement at each of the two time points.

Score Reporting. TAKS results are reported in two ways: raw score and scale scored. The raw score is the number of correct answers obtained out of the total number of answers. A scale score is a conversion of the raw score into a score that is common for

all forms of the assessment. This allows direct comparisons of student performance between different test administration and sets of test questions. Scale scores range from 1000 to 3200, with a score of 2100 marking the cut point for meeting standard performance level, and a score of 2400 marking the cut point for commended performance. In the current study, students' scale scores in math and reading were examined as measures of academic achievement.

Psychometric Properties. Reliability and validity estimates for TAKS are provided in the Student Assessment Division's Technical Digest (TEA, 2010). The KR20 was used to calculate the internal consistency reliability estimates. For the 2008-2009 school year, most TAKS internal consistency reliabilities range from the high 0.80s to low 0.90s, which are in the good to excellent range for student-level interpretations. SEM values for this same year were approximately 31 to 60 scale score points across tests and grades. Evidence for the validity of the TAKS was collected based on test content, the response process, the internal structure, relationships with other variables, and the consequences of testing. Results support the TAKS to be a valid measurement of grade-level knowledge as determined by underlying TEKS criteria (TEA, 2010). Additionally, studies conducted by Pearson (Davies, O'Malley & Wu, 2007) suggest English TAKS and Spanish TAKS are comparable in measuring the underlying constructs of math, reading, and science. For more information, consult the state's annual technical digest (TEA, 2010).

Procedure

This study was conducted in compliance with the ethical principles and standards of research set forth by the American Psychological Association and The University of Texas at Austin. Prior to beginning the study, the study was approved by the Departmental Review Committee of the Department of Education Psychology at the University of Texas at Austin, by the Institutional Review Board at The University of Texas at Austin (IRB# 2013-07-0061), and by the Texas Education Research Center Joint Advisory Board. Data security and confidentiality protocols as outlined by the Texas ERC were strictly followed. All analyses were conducted on site at Texas ERC and results were de-identified with small cells masked in all output. All statistical output released to the investigator were reviewed and approved by the Texas ERC director. A Texas ERC researcher confidentiality agreement was completed as well as researcher training on Family Educational Rights and Privacy Act (FERPA) prior to the release of data.

Statistical Analyses

Data analyses began with the computation of descriptive statistics for each district of interest and comparison group (e.g., ELL type, language programming model received) included in the study. The IBM SPSS Statistics package version 22 was employed to analyze the data using a factorial analysis of variance (two-way ANOVA) model to test the null hypotheses of no group differences by district of attendance for the first two research questions. The third research question was analyzed in two parts, first using two-way ANOVA for all student groups by district, followed by an analysis of

covariance (ANCOVA) for the ELL groups only by district with TELPAS score as the covariate variable. Prior to conducting the formal analysis of data, preliminary steps were conducted to strengthen the validity of the conclusions. Appropriate post-hoc analyses were conducted to further explore significant main effects and interactions between variables. District of student attendance was included in the statistical models as a fixed factor as all major urban districts in Texas were included in the study; therefore district was not considered a random factor. District of attendance was included as an independent variable rather than as a covariate variable after discovering that districts varied significantly in their provision of language programming and subsequent achievement results.

This study originally proposed to employ a growth curve analysis using a hierarchical linear modeling (HLM) framework (Appendix B) due to the nested structure of student data within schools. After the study's scope was broadened to include 10 major urban districts in Texas and considering available resources, the analyses were simplified to two-way ANOVA and ANCOVA for several reasons. First, by examining only 10 districts out of all of Texas, it was unlikely that estimates of variance components at the district-level would be adequate, and the number of units at the district-level would be too small to meet the minimum requirement for adequate statistical power, which precludes the use of HLM (Raudenbush & Bryk, 2002). Second, in applying conventional linear models using repeated correlated measures such as in the current study, a repeated-measure multivariate analysis of variance (MANOVA) would be the most appropriate method as it takes into consideration the correlated nature of the dependent variables

(Stevens, 2007). The current analysis, however, takes a simplified approach by running separate ANOVAs at each time point with reading and math achievement data examined individually. This decision was made to avoid a full examination of the change trajectories that would require tackling the data issue of autocorrelation of the dependent variable over time. Furthermore, the SPSS package available on the Texas ERC workstations offer limited features and do not allow for repeated measure ANOVA or MANOVA analyses.

Research Questions and Hypotheses

Research question 1. Controlling for demographics (e.g., Latino, low-SES), and compared with non-ELL peers, are group mean differences present between students enrolled in various primary language program models (ESL vs. bilingual vs. equal years ESL/bilingual vs. no services) during elementary years on ELL student academic achievement in secondary grades?

Hypothesis 1. Students who primarily received bilingual education during elementary years are hypothesized to perform significantly higher on the math and reading TAKS in 7th and 9th grades, compared to students enrolled in equal years ESL/bilingual, ESL, and those who did not receive any language programming, yet not as high as the non-ELL comparison population. Table 3 ranks the hypothesized order of achievement by classroom model.

Table 3

Research Question 1, Hypothesis 1

Hypothesized Achievement Outcome Rank	Classroom Model
1.	Non-ELL Comparison
2.	Bilingual
3.	Equal ESL/bilingual
4.	ESL
5.	No Services

Rationale. Thomas and Collier (2002) identified students enrolled in dual language bilingual programs as having the highest levels of long-term achievement, with students enrolled in transitional and developmental bilingual programming having the next highest levels of achievement, followed by students in ESL. Based on this study, students enrolled in dual-language programming have similar achievement outcomes as their non-ELL peers, whereas students enrolled in developmental and transitional bilingual programming attain on average below the 50th percentile (Thomas & Collier, 2002). However, because only 10% of Texas school districts report the implementation of dual-language programs (Lara-Alecio, Galloway, Irby, Rodriquez, & Gomez, 2004), this study is unable to examine the subcategories of bilingual enrollment individually, and hypothesizes that students enrolled in bilingual education, as a whole, will have higher achievement than students in ESL programming. Students who received equal years of ESL and bilingual are hypothesized to have higher achievement than students who received ESL programming alone, as exposure to some bilingual education may still be beneficial. ELL students who did not receive any language programming are hypothesized to have the lowest relative achievement. Finally, ELL students in general

are posited to not achieve as highly as their non-ELL peers, regardless of classroom model received.

Research question 2. How do former-ELLs (defined as reclassified to English proficient in 5th grade or earlier) perform on measures of academic achievement in 7th and 9th grade compared to long-term ELLs (defined as reclassified to English proficient after 5th grade, or not reclassified by end of study) and their non-ELL counterparts?

Hypothesis 2. As a group, long-term ELL students will have significantly lower scores on state level assessments (TAKS) in math and reading compared to the former-ELL group at both time points. Both groups will have significantly lower scores at both time points than their non-ELL peers. Table 4 ranks the hypothesized order of achievement by ELL group.

Table 4

Research Question 2, Hypothesis 2

Hypothesized Achievement Outcome Rank	Group
1.	Non-ELLs
2.	Former-ELLs
3.	Long-term ELLs

Rationale. Based on the existing literature on the low academic achievement of long-term ELLs (Menken & Kleyn, 2009; Olsen, 2010; Slama, 2012), it is predicted that these students are at a unique disadvantage compared to their former-ELL peers who were able to achieve English proficiency at a relatively young age. Prior research is inconsistent about the performance of former-ELLs compared to their non-ELL peers, with some studies suggesting higher academic achievement (Ardasheva, Tretter, &

Kinny, 2012; Kim, Herman, & National Center for Research on Evaluation, 2009), while other studies suggest that former-ELLs continue to be at an academic disadvantage post-reclassification (de Jong, 2004, Thomas & Collier, 2002). The current hypothesis that former-ELLs will not score as highly as non-ELLs is based on the research that former-ELLs who perform better than non-ELLs tend to be high-achievers who exit language programming in three or less years (Flores, Batalova, & Fix, 2012), which is only a subset of this study's former-ELL sample. Furthermore, considering the challenge of attaining both English language proficiency and content knowledge simultaneously in order to catch up to their non-ELL counterparts, it is predicted that there will still exist an achievement gap between former-ELL students and their non-ELL peers.

Research question 3. Assuming that hypothesis 2 is confirmed and former-ELLs have higher achievement outcomes than long-term ELLs in 7th and 9th grade, what individual differences between groups significantly contribute to differential achievement outcomes?

Hypothesis 3. This is an exploratory analysis examining if significant mean differences in achievement between the two ELL groups exist across several independent variables at both time points. Hence, there are no directional hypotheses associated with this research question. Rather, it is an open-ended examination of various factors between the groups including: primary classroom model received, district of enrollment, and immigrant status. This research question was examined in two parts: first, comparing the two groups of ELL students to their non-ELL counterpart, second, looking at only the target ELL groups with baseline TELPAS score as a covariate variable. These analyses

also serve to explore group differences delineated by the first two research questions by examining possible interactions between independent variables.

Rationale. The inclusion of the selected independent variables is based on existing literature that suggests these are relevant predictors of ELL achievement. Assuming that Hypothesis 2 is correct, it will be important to examine the factors that may contribute to why some ELLs are able to gain English proficiency, while others struggle with long-term ELL status. Previous research has consistently identified bilingual education as a predictor of academic achievement in ELLs (Collier 1987, Cummins 1979; Goldenberg, 2008; Thomas & Collier, 1997, 2002). Research has also suggested that stronger English proficiency levels at an earlier age to be a predictor of achievement (Halle et al., 2012; Kim, Herman & National Center for Research on Evaluation, 2012). By controlling for baseline English proficiency, it will be possible to examine if group differences in achievement continue to exist between former and long-term ELLs. Finally, Slama (2012) found that students who are U.S. born are more likely to perform lower on achievement measures and to become long-term ELLs than their immigrant counterparts.

Chapter Four: Results

This chapter aims to describe the cohort of students in this study and the relationships between their academic achievement and key study variables. The variables included primary classroom language programming model received, ELL status, immigrant status, district of attendance, and baseline English language proficiency.

Descriptive Statistics

District attrition. Students included in the study remained in one of the 10 target school districts for the duration of the study. The starting sample of all students enrolled in 3rd grade within target districts during the 2003-2004 school year was 69,107 and the ending cohort sample was 51,263 during the 2009-2010 school year. Approximately 26% (N = 17,844) of all students were lost across study years. After limiting sample to students who attended only the major urban districts for the duration of the study, the sample size was 38,114. From this group, the number of students who met study demographic criteria of Latino, free or reduced lunch (low-income), and Spanish and/or English speaking was 21,435.

Special education. Students who received special education at any point during the study were removed from the sample of 21,435, yielding a final sample size of 18,188. While students in special education were not included in the study's analyses, descriptive statistics were run prior to removing them from the dataset in hopes of better understanding the student population as a whole. Of all students from the total sample of 21,435 who were non-ELL (N = 7312), 19.7% (N = 1,437) received special education during the study years. Of all former-ELL students (N = 9,327), 4.7% (N = 440) received

special education. Finally, of all long-term ELLs (N = 4,796), 28.6% (N = 1,370) received special education. In sum, 15.1% (N = 3,247) of students received special education at any point during the study years of 3rd through 9th grade and were removed from the full sample of 21,435 to yield a final study sample size of 18,188. Notably, there were 172 students who were enrolled in special education in 2nd grade but were no longer receiving these services during the study years. These students remained in the final sample.

ELL type and grade of exit. Descriptive statistics for ELL group enrollment are reported in Table 5. Grade of exit from language programming descriptive statistics are reported in Table 6. The shaded cells of Table 6 denote long-term ELLs, while the cells above that are not shaded are former ELLs.

Table 5

ELL Type

Group	Frequency	Percent
Former-ELL	8887	48.9
Long Term-ELL	3426	18.8
Non-ELL	5875	32.3
Total Sample	18188	100

Table 6

Grade Exited Language Programming for ELL Students

Grade	Frequency	Percent
Never enrolled	811	6.6
2nd Grade	773	6.3
3rd Grade	3208	26.1
4th Grade	2087	16.9
5th Grade	2136	17.3
6th Grade	734	6.0
7th Grade	475	3.9
8th Grade	467	3.8
Not Exited by end of 9th Grade	1622	13.2
Total	12313	100

Note. Shaded cells denote long-term ELL students, not shaded cells above are former-ELLs.

Table 6 shows that the vast majority of students who leave ELL programming do so by the end of 5th grade. On the other hand, 46.8% (N = 1,622) of all long-term ELLs do not exit programming by the end of 9th grade. It is worth noting that there were approximately 6% (N = 773) of students who were ELL in 2nd grade but no longer ELL starting in 3rd grade and through the study years. As explained in chapter three, the available data required deriving the comparison non-ELL student group by parsing out students who were not ELL in 2nd grade, the first year of data availability. Students who were ELL at any time between 2nd and 9th grade were considered part of the ELL student groups.

6.6% (N = 811) of identified ELL students never receive any type of language services. Table 7 explores the possible reasons for their lack of language programming through examining the LEP permissions for this group of students. The majority (N =

502, 61.9%) of these students did not receive programming because parents denied specialized language services.

Table 7

LEP Permissions for Students Never Enrolled in Language Programming

Reason	Frequency	Percent
Parents denied programming	502	61.9
Missing LEP permission data	234	28.9
Parent not contacted	34	4.2
Total	811	100

Primary language programming received. Table 8 outlines the primary language programming classroom model that ELL students received during elementary years of data availability, from 2nd through 5th grade. The "No Services" category (*N* = 831, 6.7%) is comprised of students who did not receive services at any point during their elementary years. Notably, this number differs slightly from the students in Table 7 who did not receive language programming at any point during their education through 9th grade. 25% (*N* = 3,127) of ELL students received a combination of both bilingual and ESL instruction during their education; however, only 10.3% (*N* = 1270) of ELLs received exactly equal years of the two classroom models during elementary school. The majority (56.6%) of ELL students are enrolled in bilingual programming during elementary years.

Table 8

Primary Language Programming Received in Elementary

Programming Type	Frequency	Percent
ESL	3247	26.4
Bilingual	6965	56.6
Equal ESL/Bilingual	1270	10.3
No Services	831	6.7
Total	12313	100

TELPAS scores. This measure of baseline English language proficiency was used as a covariate variable in research question 3. Table 9 outlines the distribution of composite TELPAS scores for ELL students in 3rd grade. 13% (*N* = 1,596) of ELLs did not have a valid TELPAS score. In order to exit ELL designation, students must achieve a TELPAS score of 3 (advanced) or higher and be enrolled in U.S. education for at least three years.

Table 9

TELPAS Score at Baseline

Descriptor	Score	Frequency	Percent
Beginning	1.00	596	4.8
	1.50	812	6.6
Intermediate	2.00	1461	11.9
	2.50	2022	16.4
Advanced	3.00	2575	20.9
	3.50	2286	18.6
Advanced High	4.00	965	7.8
	Missing	1596	13
	Total	12313	100

District of enrollment. Originally, this study did not conceptualize district of enrollment as a significant independent variable. Therefore, descriptive statistics did not

examine district differences for the complete student sample. Tables 10 and 11 utilize the student sample associated with 7th grade reading TAKS scores. This sample (N = 16,168) is smaller than the total sample size (N = 18,188), which is attributable to missing data from students who did not take the test on the first administration date.

Table 10

District of Enrollment by Primary Language Programming Received

District	Language Programming	Frequency	Percent of District	Percent of Total
1	No Services	9	1.4	*
	Elem. ESL	215	34.3	1.3
	Elem. Bilingual	174	27.8	1.1
	Elem. Equal ESL/Bilingual	82	13.1	0.6
	Non-ELL	147	23.4	0.9
	Total	627	100	3.9
2	No Services	32	3.2	0.2
	Elem. ESL	47	4.7	0.3
	Elem. Bilingual	594	58.9	3.7
	Elem. Equal ESL/Bilingual	21	2.1	0.1
	Non-ELL	315	31.2	1.9
	Total	1009	100	6.2
3	No Services	134	4	0.8
	Elem. ESL	1738	51.6	10.7
	Elem. Bilingual	375	11.1	2.3
	Elem. Equal ESL/Bilingual	652	19.3	4.0
	Non-ELL	472	14	2.9
	Total	3371	100	20.8
4	No Services	183	10.7	1.1
	Elem. ESL	<5	*	*
	Elem. Bilingual	938	54.8	5.8
	Elem. Equal ESL/Bilingual	<5	*	*
	Non-ELL	590	34.4	3.6
	Total	1713	100	10.6
5	No Services	21	1.3	0.1
	Elem. ESL	628	39.4	3.9
	Elem. Bilingual	392	24.6	2.4
	Elem. Equal ESL/Bilingual	274	17.2	1.7
	Non-ELL	277	17.4	1.7
	Total	1592	100	9.8
6	No Services	164	4.7	1.0
	Elem. ESL	195	5.5	1.2
	Elem. Bilingual	2177	61.8	13.5
	Elem. Equal ESL/Bilingual	79	2.2	0.5
	Non-ELL	905	25.7	5.6
	Total	3520	100	21.8

Table 10 continued

7	No Services	28	5.7	0.2
	Elem. ESL	<5	*	*
	Elem. Bilingual	130	26.7	0.8
	Elem. Equal ESL/Bilingual	<5	*	*
	Non-ELL	326	66.9	2.0
	Total	487	100	3.0
8	No Services	63	6.5	0.4
	Elem. ESL	<5	*	*
	Elem. Bilingual	173	17.8	1.1
	Elem. Equal ESL/Bilingual	<5	*	*
	Non-ELL	733	75.6	4.5
	Total	970	100	6.0
9	No Services	73	4.7	0.5
	Elem. ESL	<5	*	*
	Elem. Bilingual	481	31.2	3.0
	Elem. Equal ESL/Bilingual	<5	*	*
	Non-ELL	989	64.1	6.1
	Total	1543	100	9.5
10	No Services	60	4.5	0.4
	Elem. ESL	<5	*	*
	Elem. Bilingual	646	48.4	4.0
	Elem. Equal ESL/Bilingual	<5	*	*
	Non-ELL	629	47	3.9
	Total	1336	100	8.3
Total	No Services	767	-	4.7
	Elem. ESL	2829	-	17.5
	Elem. Bilingual	6080	-	37.6
	Elem. Equal ESL/Bilingual	1109	-	6.9
	Non-ELL	5383	-	33.3
	Total	16168	-	100

Note. Small cell<5 masked, as well as corresponding cell in same district. Associated percentages are marked with * as these were not possible to calculate.

Table 10 showcases the primary language programming received during elementary years for each school district. Although this sample is a subset of the study's total sample, there are noticeable differences between primary language programming received throughout the study's duration (Table 8) and the primary programming received during elementary years. 38.5% (N = 4,743) of all ELLs primarily received ESL, while 26.2% (N = 2,829) of this sample primarily received ESL during elementary school. 43% (N = 5,291) of ELLs primarily received bilingual, while 55.9% (N = 6,080) of ELLs primarily received bilingual in elementary years. These differences illustrate that

ELLs move away from bilingual education towards ESL programming as they enter secondary education.

There exists significant variation between districts on the provision of primary language programming. While some districts appear to utilize both ESL and bilingual education comparably (e.g., district 1, district 5), others appear to primarily provide bilingual (e.g., district 4, district 7, district 8, district 9, district 10) or ESL (e.g., district 3). Notably, the two largest urban districts, district 3 and district 6, which together comprise over 40% of this study's sample, utilize different primary language program models, with 51.6% of district 3's students enrolled in ESL, and 61.8% of district 6's students in bilingual.

ELL type by school district is outlined in Table 11. Similar to language programming, there are noticeable differences between districts in their makeup of each ELL group. Half of the districts have former-ELLs as the dominant group (district 1, district 3, district 4, district 5, district 6), while three districts have non-ELLs comprising the majority of their student population (district 7, district 8, district 9). District 2 is the only district with relatively equal distribution of students between all ELL types. It is important to keep in mind the differences in student composition when interpreting differential achievement results by district.

Table 11

District of Enrollment by ELL Type

District	ELL Group	Frequency	Percent of District	Percent of Total
1	Non-ELL	147	23.4	0.9
	Former-ELL	345	55	2.1
	Long Term-ELL	135	21.5	0.8
	Total	627	100	3.9
2	Non-ELL	315	31.2	1.9
	Former-ELL	380	37.7	2.4
	Long Term-ELL	314	31.1	1.9
	Total	1009	100	6.2
3	Non-ELL	472	14	2.9
	Former-ELL	2459	72.9	15.2
	Long Term-ELL	440	13.1	2.7
	Total	3371	100	20.8
4	Non-ELL	590	34.4	3.6
	Former-ELL	842	49.2	5.2
	Long Term-ELL	281	16.4	1.7
	Total	1713	100	10.6
5	Non-ELL	277	17.4	1.7
	Former-ELL	1174	73.7	7.3
	Long Term-ELL	141	8.9	0.9
	Total	1592	100	9.8
6	Non-ELL	905	25.7	5.6
	Former-ELL	2042	58	12.6
	Long Term-ELL	573	16.3	3.5
	Total	3520	100	21.8
7	Non-ELL	326	66.9	2.0
	Former-ELL	110	20.5	0.7
	Long Term-ELL	51	10.5	0.3
	Total	487	100	3.0
8	Non-ELL	733	75.6	4.5
	Former-ELL	188	19.4	1.2
	Long Term-ELL	49	5	0.3
	Total	970	100	6.0
9	Non-ELL	989	64.1	6.1
	Former-ELL	430	27.9	2.7
	Long Term-ELL	124	8	0.8
	Total	1543	100	9.5
10	Non-ELL	629	47.1	3.9
	Former-ELL	310	23.2	1.9
	Long Term-ELL	397	29.7	2.4
	Total	1336	100	8.3
Total	Non-ELL	5383	-	33.3
	Former-ELL	8280	-	51.2
	Long Term-ELL	2505	-	15.5
	Total	16168	-	100

Missing Data

The final sample of students who met study criteria is 18,188. Not all students had valid data on all study variables. Listwise deletion was employed to address the problem of missing data. If a case had missing data for any of the variables included in the analysis, this case was simply excluded from the analysis. Listwise deletion assumes data are missing completely at random (MCAR). The data on *Y* are said to be MCAR if the probability of missing data on *Y* is unrelated to the value of *Y* itself or to the values of any other variables in the data set (Allison, 2002). MCAR is a strong assumption, and does not hold in the current data set. A weaker assumption is that the data are missing at random (MAR). The data on *Y* are said to be MAR if the probability of missing data on *Y* is unrelated to the value of *Y*, after controlling for other variables in the analysis (Allison, 2002). There are no definitive tests available to gauge whether the MCAR or MAR condition is satisfied.

Statistical approaches exist for a more robust and efficient treatment of non-ignorable missing data such as: the full information maximum likelihood (FIML) estimator (Agresti & Finlay, 1997), the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977), and the multiple imputation method (Rubin, 1987). The current study did not employ any of the above statistical approaches to address missing data because the proportion of students who had missing data in this study is relatively small. The final analyzed sample for 7th grade reading scores is comprised of 16,168 students, which means 11.1% (2,020 cases) of the original sample was excluded using listwise deletion. For 7th grade math, the proportion excluded is also 11.1% (2,023

cases). From 7th to 9th grade, the proportion excluded for reading scores was 4.6% (744 cases), and for math it was 4.8% (781 cases). The proportion of cases missing from the total sample for 9th grade reading was 15.2% (2,764 cases), and for 9th grade math it was 15.3% (2,798 cases).

Analyses were conducted to examine the pattern of missing data. Based on available non-missing variables, a series of chi-square tests was conducted using the dichotomous variable of "sample analyzed vs. sample missing" as one variable, and each of the following non-missing categorical variables as the other: gender, free or reduced lunch, district, and ELL status, all at baseline. These analyses were conducted with both math and reading 7th grade samples to examine the pattern of missing data from the original sample. Results show that the missing sample was significantly different from the original sample across all variables. For 7th grade reading outcomes, missing cases were more likely to be ELL students, X^2 (1, N = 18,188) = 115.79, $p < .01$; male, X^2 (1, N = 18,188) = 33.35, $p < .01$; and receive free lunch as opposed to reduced price lunch or "other" lunch status, X^2 (2, N = 18,188) = 38.66, $p < .01$. There were also significant district differences in the missing sample compared to the original sample, X^2 (9, N = 18,188) = 206.31, $p < .01$. Similar results were found for 7th grade math outcomes, with missing cases more likely to be ELL students, X^2 (1, N = 18,188) = 116.82, $p < .01$; male, X^2 (1, N = 18,188) = 35.31, $p < .01$; and receive free lunch, X^2 (2, N = 18,188) = 38.16, $p < .01$. District differences were found as well, X^2 (9, N = 18,188) = 201.50, $p < .01$.

In order to examine the pattern of missing data between 7th and 9th grade study points, additional independent samples t-tests were conducted using the dichotomous

variable of "sample analyzed at 9th vs. sample missing from 7th to 9th" as one variable, and 7th grade achievement outcomes in both math and reading as the other variable. Results indicate that missing cases from 7th to 9th grade (M = 2135.09, SD = 178.79) scored significantly lower on the 7th grade reading TAKS than non-missing cases (M = 2244.99, SD = 158.83), $t(16166) = -19.46$, $p = .00$. A similar pattern emerged for math, as missing cases from 7th to 9th grade (M = 2095.12, SD = 155.67) scored significantly lower on the 7th grade math TAKS than non-missing cases (M = 2207.96, SD = 161.08), $t(16169) = -20.28$, $p = .00$. Limitations of using listwise deletion under these non-MCAR conditions are addressed in the discussion chapter.

Tests of Research Questions

Three multi-part research hypotheses were stated and tested. Factorial analysis of variance (two-way ANOVA) was used to test research questions 1, 2, and the first part of research question 3. Analysis of covariance (ANCOVA) was used to test the second part of research question 3. Each research question utilized four separate ANOVA/ANCOVA analyses, with grade and subject of achievement scores (7th, 9th; reading, math) serving as the dependent variables. Additional follow-up tests were conducted to examine significant interactions between independent variables, when appropriate. Post-hoc analyses using Tukey Honest Significant Differences Test (Tukey HSD) were utilized to compare group mean differences. Effect sizes were not reported, as the majority of the partial eta squared effect sizes were very small (<.01); however, having sufficiently large sample sizes still yielded significant results.

In conducting these analyses, I recognize that the statistical inference assumption of independence associated with factorial ANOVA was violated, as students are not educated individually but are rather nested in classrooms, schools, and districts. As a result, the Type 1 error rate of incorrectly rejecting a true null hypothesis is increased, that is, the likelihood of finding significant differences when differences do not truly exist.

Analyses and Results for Hypothesis 1

Students who primarily received bilingual education during elementary years are hypothesized to perform significantly higher on the math and reading TAKS in 7th and 9th grades, compared to students enrolled in equal years ESL/bilingual, ESL, and those who did not receive any language programming, yet not as high as the non-ELL comparison population.

For this research question, the focus of the ANOVAs was to determine whether the language-programming factor (ESL, bilingual, equal ESL/bilingual, no services, comparison non-ELL), and district of enrollment factors (10 districts) account for a potentially meaningful amount of variance in TAKS achievement scores. Although there were significant interaction effects between the two factors of district and language programming in each ANOVA, post-hoc analyses to examine these interactions were not conducted. It is not possible to extrapolate from the available data what it means in theory and practice that there is an interaction between district and language programming beyond the existence of variability in local practice.

Descriptive statistics. Table 12 provides the sample size, mean score, and standard deviation for each language program factor, across the four analyses. Note that a TAKS score of 2100 denotes passing, and scores of 2400 and above are required for commended performance. Figure 3 provides a graphical representation of this data.

Table 12

Mean Scores by Primary Language Programming Received in Elementary School

Language Programming	*N*	*M*	*SD*
7th Grade Reading			
No Services	767	2265.10	171.25
ESL	2829	2234.99	163.14
Bilingual	6080	2216.45	158.95
Equal ESL/Bilingual	1109	2205.23	153.62
Non-ELL Comparison	5383	2270.54	158.65
7th Grade Math			
No Services	762	2222.78	171.84
ESL	2828	2192.46	166.46
Bilingual	6084	2202.48	163.04
Equal ESL/Bilingual	1109	2176.32	157.54
Non-ELL Comparison	5388	2208.21	159.54
9th Grade Reading			
No Services	723	2285.60	139.23
ESL	2685	2260.75	135.64
Bilingual	5804	2254.36	133.21
Equal ESL/Bilingual	1064	2241.16	128.75
Non-ELL Comparison	5148	2292.78	131.69
9th Grade Math			
No Services	715	2247.69	224.17
ESL	2675	2214.87	214.47
Bilingual	5802	2227.91	218.39
Equal ESL/Bilingual	1062	2194.80	215.38
Non-ELL Comparison	5136	2221.04	212.09

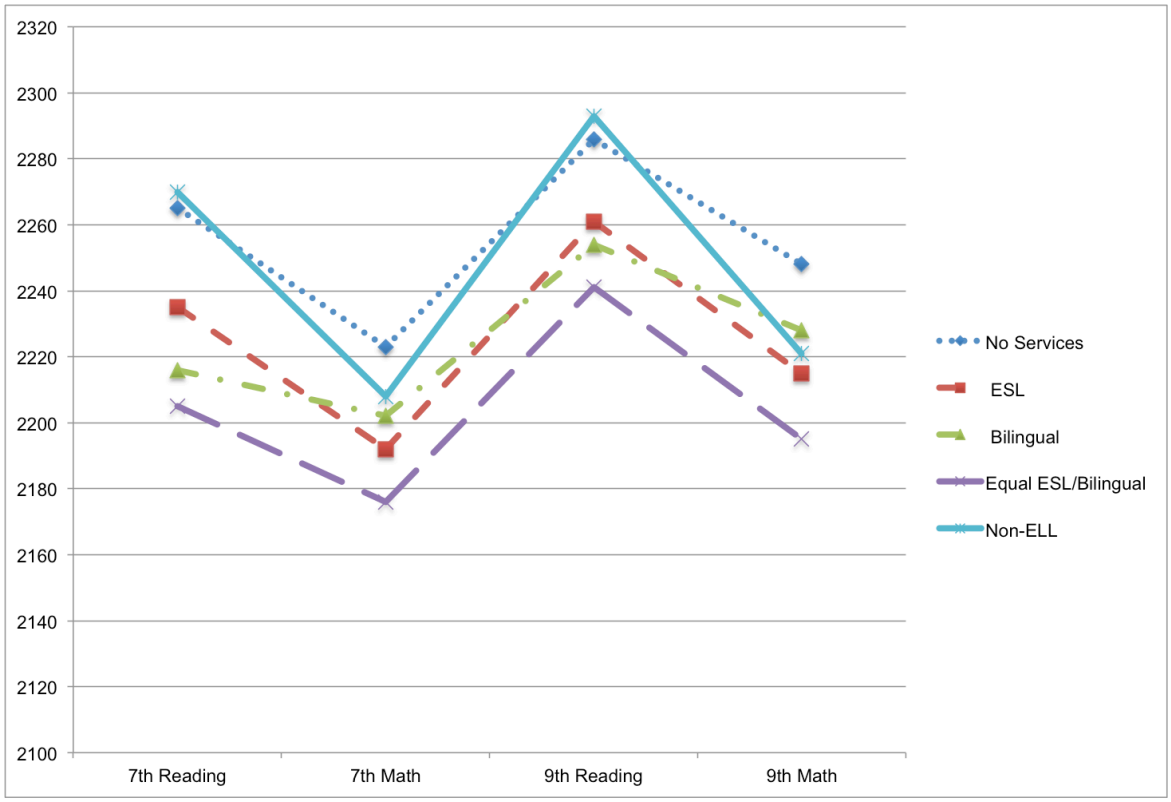

Figure 3. Mean achievement scores of each language program model at all study points.

7th grade reading. Prior to conducting the formal analysis of the data, Levene's test for the equality of variance was conducted. Results indicate that population cell variances are unequal at the .05 level, $F(44, 16123) = 2.1$, $p = .00$. A stricter alpha of .01 was utilized to control for the possible inflation of the Type 1 error rate. An ANOVA for language programming by district was conducted using 7th grade reading scores. Tables 13 summarizes the significant results related to this analysis, and shows that there were significant differences between the four language program models and the comparison group, $F(4, 16123) = 48.32$, $p < .01$. Significant differences also exist between the 10 school districts, $F(9, 16123) = 3.27$, $p < .01$. District results are further discussed separately below in its own section.

Table 13

7th Grade Reading Scores by Language Programming and District (N = 16,168)

Source	*df*	*F*	Sum of Squares	Mean Square
Language Program	4	48.32*	4846900.24	1211725.06
District	9	3.27*	736899.18	81877.69
Lang. Prog. x District	36	5.40*	4196723.60	135378.18
Error	16123		404361294.43	(25079.78)

Note: $*p < .01$

A post-hoc Tukey's HSD analysis was conducted to examine the relative performance of each type of language program model. Results show that non-ELLs were the highest performing group (M = 2270.54), followed by students who received no services (M = 2265.10). There were no significant differences between these two groups. However, these two groups performed significantly higher than the other groups ($p <$ .05). Primary ESL (M = 2234.99) was the third highest performing language model, with a mean score that was significantly higher than the lowest two performing groups of primary bilingual (M = 2216.45) and equal ESL/bilingual (M = 2205.23). Table 14 illustrates the differences in observed group means between the various language programs.

Table 14

Differences in Means for 7th Grade Reading Scores by Language Programming

Contrast	Estimate	*SE*	95% Confidence Interval
Bilingual vs. ESL	-18.54*	3.60	-28.37, -8.71
Bilingual vs. No Svcs.	-48.65*	6.07	-65.21, -32.10
Bilingual vs. Equal ESL/Bil.	11.22	5.17	-2.88, 25.33
Bilingual vs. Non-ELL	-54.09*	2.96	-62.18, -46.00
ESL vs. No Svcs.	-30.11*	6.45	-47.70, -12.52
ESL vs. Equal ESL/Bil.	29.76*	5.61	14.46, 45.07
ESL vs. Non-ELL	-35.55*	3.68	-45.58, -25.52
No Svcs. vs. Equal ESL/Bil	59.88*	7.44	39.59, 80.17
No Svcs. vs. Non-ELL	-5.44	6.11	-22.11, 11.24
Non-ELL vs. Equal ESL/Bil.	65.31*	5.22	51.07,79.56

Note: *Significant at the .05 level using the Tukey HSD procedure.

7th grade math. Similar to the previous analysis, Levene's test for the equality of variance results indicated unequal population cell variances at the .05 level, $F(44, 16126) = 2.1$, $p = .00$. A stricter alpha of .01 was utilized to control for the possible inflation of the Type 1 error rate. An ANOVA for language programming by district was conducted using 7th grade math scores. Tables 15 summarizes results related to this analysis, and shows that there were no significant differences for the overall effect of language programming type, $F(4, 16126) = 2.91$, $p = .02$. However, there exist significant differences in 7th grade math scores between the 10 school districts, $F(9, 16126) = 7.59$, $p < .01$. Post-hoc analyses for language programming were not conducted.

Table 15

7th Grade Math Scores by Language Programming and District (N = 16,171)

Source	*df*	*F*	Sum of Squares	Mean Square
Language Program	4	2.91	300518.17	75129.54
District	9	7.59*	1766389.41	196265.49
Lang. Prog. x District	36	3.12*	2497430.45	80562.27
Error	16126		417120462.89	(25866.33)

Note: $*p < .01$

9th grade reading. Again, Levene's test for the equality of variance results indicated unequal population cell variances at the .05 level, $F(44, 15379) = 1.98, p = .00$. A stricter alpha of .01 was similarly used in this analysis to control for the possible inflation of the Type 1 error rate. An ANOVA for language programming by district was conducted using 9th grade reading scores. Tables 16 summarizes the results related to this analysis, and shows that there were significant differences between the four language program models and the comparison group, $F(4, 15379) = 33.89, p < .01$. There were no significant group mean differences associated with district in this analysis, $F(9, 15379) = 1.90, p = .05$.

Table 16

9th Grade Reading Scores by Language Programming and District (N = 15,424)

Source	*df*	*F*	Sum of Squares	Mean Square
Language Program	4	33.89*	2381969.412	595492.35
District	9	1.90	300382.46	33375.83
Lang. Prog. x District	36	3.64*	1983558.58	63985.76
Error	15379		270233462.19	(17571.59)

Note: $*p < .01$

A post-hoc Tukey's HSD analysis was conducted to examine the relative performance of each type of language program model. Results show that non-ELLs were

the highest performing group (M = 2292.78), followed by students who received no services (M = 2285.60). There were no significant differences between these two groups. These two groups performed significantly higher than the other groups ($p < .05$). Primary ESL (M = 2260.75) was the third highest performing language model, followed by primary bilingual (M = 2254.36). There were no significant differences between these two groups. The lowest performing group was equal ESL and bilingual (M = 2241.16), which performed significantly lower than all other language programming groups. Table 17 illustrates the differences in observed means between the various language programs.

Table 17

Differences in Means for 9th Grade Reading Scores by Language Programming

Contrast	Estimate	*SE*	95% Confidence Interval
Bilingual vs. ESL	-6.39	3.09	-14.83, 2.05
Bilingual vs. No Svcs.	-31.24*	5.23	-45.50, -16.98
Bilingual vs. Equal ESL/Bil.	13.20*	4.42	1.14, 25.26
Bilingual vs. Non-ELL	-38.42*	2.54	-45.34, -31.50
ESL vs. No Svcs.	-24.85*	5.55	-40.00, -9.70
ESL vs. Equal ESL/Bil.	19.59*	4.80	6.49, 32.69
ESL vs. Non-ELL	-32.03*	3.16	-40.64, -23.42
No Svcs. vs. Equal ESL/Bil	44.44*	6.39	27.01, 61.87
No Svcs. vs. Non-ELL	-7.18	5.27	-21.54, 7.18
Non-ELL vs. Equal ESL/Bil.	51.62*	4.46	39.44, 63.80

Note: *Significant at the .05 level using the Tukey HSD procedure.

9th grade math. As with the prior analyses, Levene's test for the equality of variance results indicated unequal population cell variances at the .05 level, $F(44, 15345) = 1.6$, $p = .01$, and a stricter alpha of .01 was used. An ANOVA for language programming by district was conducted using 9th grade math scores. Tables 18 summarizes the results related to this analysis, and shows that similar to 7th grade math scores, there were no significant differences between the language program models, $F(4,$

15345) = 1.93, p = .10. District of enrollment was indeed significant, $F(9, 15345) = 5.10$, $p < .01$. Post-hoc analyses for language program were not completed.

Table 18

9th Grade Math Scores by Language Programming and District (N = 15,390)

Source	*df*	*F*	Sum of Squares	Mean Square
Language Program	4	1.93	352934.03	88233.51
District	9	5.10*	2097597.42	233066.38
Lang. Prog. x District	36	2.76*	3911356.91	126172.80
Error	15345		701669585.73	(45726.27)

Note: *$p < .01$

Summary. It appears that there are significant group differences between ELLs who received different primary language programming in elementary years on reading scores during secondary years, but not on math scores. The hypothesized outcome ranks outlined in Table 3 were not congruent with actual results; therefore hypothesis 1 was not confirmed. Table 19 illustrates the actual outcome ranks of language program models. This table portrays a tie between no services and non-ELLs, as there were no significant differences between these scores at both time points, with no services performing the highest on 7th grade reading, while non-ELLs took the top rank when looking at 9th grade reading scores. The remaining language program models followed the same rank order in post-hoc analyses examining both 7th and 9th grade reading scores.

Table 19

Research Question 1 Results

Actual Achievement Outcome Rank	Classroom Model
1.	Non-ELL Comparison
1.	No Services
3.	ESL
4.	Bilingual
5.	Equal ESL/Bilingual

Analyses and Results for Hypothesis 2

As a group, long-term ELL students will have significantly lower scores on state level assessments (TAKS) in math and reading compared to the former-ELL group at both time points. Both groups will have significantly lower scores at both time points than their non-ELL peers.

For this research question, the focus of the ANOVAs was to determine whether the ELL type factor (former ELL, long-term ELL, non-ELL), and district of enrollment factors (10 districts) account for a potentially meaningful amount of variance in TAKS achievement scores. Similar to the previous research question, significant interactions between district and ELL type were present in every analysis, but were not further explored. A summary of the district factor will be reported in the following section.

Descriptive statistics. Table 20 provides the sample size, mean score, and standard deviation for each ELL group, across the four analyses, while figure 4 provides a graphical representation of this data.

Table 20

Mean Scores by ELL Type

ELL Type	*N*	*M*	*SD*
	7th Grade Reading		
Former ELL	8280	2256.53	149.67
Long-Term ELL	2505	2114.83	149.24
Non-ELL	5383	2270.54	158.65
	7th Grade Math		
Former ELL	8279	2221.21	162.98
Long-Term ELL	2504	2123.81	145.64
Non-ELL	5388	2208.21	159.54
	9th Grade Reading		
Former ELL	7936	2279.01	128.56
Long-Term ELL	2340	2181.73	125.21
Non-ELL	5148	2292.78	131.70
	9th Grade Math		
Former ELL	7908	2249.29	219.43
Long-Term ELL	2346	2132.02	185.59
Non-ELL	5136	2221.04	212.09

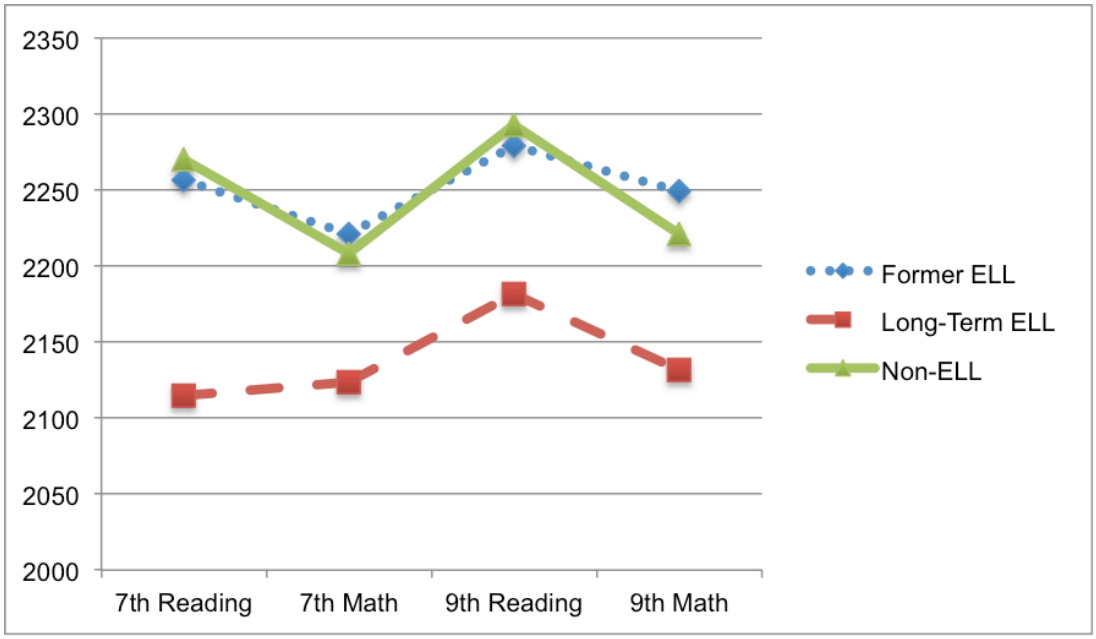

Figure 4. Mean achievement scores of each ELL type at all study points.

7th grade reading. Prior to conducting the formal analysis of the data, Levene's test for the equality of variance was conducted. Results indicate that population cell variances are unequal at the .05 level, $F(29, 16138) = 3.6$, $p = .00$. A stricter alpha of .01 was utilized to control for the possible inflation of the Type 1 error rate. An ANOVA for ELL type by district was conducted using 7th grade reading scores. Tables 21 summarizes the significant results related to this analysis, and shows that there were significant differences between the three ELL types, $F(2, 16138) = 621.63$, $p < .01$. Significant differences also exist between the 10 school districts, $F(9, 16138) = 22.46$, $p < .01$.

Table 21

7th Grade Reading Scores by ELL Type and District (N = 16,168)

Source	*df*	*F*	Sum of Squares	Mean Square
ELL Type	2	621.63*	28497353.61	14248676.81
District	9	22.46*	4632746.08	514749.57
ELL Type x District	18	7.10*	2930254.71	162791.93
Error	16138		369907028.84	(22921.49)

Note: $*p < .01$

A post-hoc Tukey's HSD analysis was conducted to examine the relative performance of each ELL group. Results show that non-ELLs were the highest performing group ($M = 2270.54$), followed by former-ELLs ($M = 2256.53$). Long-term ELLs performed the poorest ($M = 2114.83$). Each group's means were significantly different from the others at $p < .05$. Table 22 illustrates the differences in observed means between the ELL types.

Table 22

Differences in Means for 7th Grade Reading Scores by ELL Type

Contrast	Estimate	*SE*	95% Confidence Interval
Former ELL vs. Long ELL	141.70*	3.45	133.61, 149.79
Former ELL vs. Non-ELL	-14.01*	2.65	-20.22, -7.80
Non-ELL vs. Long-ELL	155.71*	3.66	147.13, 164.29

Note: *Significant at the .05 level using the Tukey HSD procedure.

7th grade math. Again, Levene's test for the equality of variance results indicated unequal population cell variances at the .05 level, $F(29, 16141) = 5.3$, $p = .00$. A stricter alpha of .01 was utilized to control for the possible inflation of the Type 1 error rate. An ANOVA for ELL type by district was conducted using 7th grade math scores. Tables 23 summarizes results related to this analysis, and shows that there were significant differences between the ELL groups, $F(2, 16141) = 241.47$, $p < .01$ and district of attendance, $F(9, 16141) = 33.64$, $p < .01$.

Table 23

7th Grade Math Scores by ELL Type and District (N = 16,171)

Source	*df*	*F*	Sum of Squares	Mean Square
ELL Type	2	241.47*	11933973.67	5966986.83
District	9	33.64*	7481833.32	831314.81
ELL Type x District	18	5.57*	2478133.66	137674.09
Error	16141		398862963.47	(24711.17)

Note: *$p < .01$

A post-hoc Tukey's HSD analysis was conducted to examine the relative performance of each ELL group. Results show that former-ELLs were the highest performing group ($M = 2221.21$) followed by non-ELLs ($M = 2208.21$). Long-term ELLs again performed the poorest ($M = 2123.81$). Each group's means were significantly

different from the others at $p < .05$. Table 24 illustrates the differences in 7th grade math score means between the ELL types.

Table 24

Differences in Means for 7th Grade Math Scores by ELL Type

Contrast	Estimate	*SE*	95% Confidence Interval
Former ELL vs. Long ELL	97.41*	3.56	89.00, 105.81
Former ELL vs. Non-ELL	13.00*	2.75	6.55, 19.45
Non-ELL vs. Long-ELL	84.41*	3.80	75.50, 93.32

Note: *Significant at the .05 level using the Tukey HSD procedure.

9th grade reading. Similar to previous analyses, Levene's test for the equality of variance results indicated unequal population cell variances at the .05 level, $F(29, 15394) = 2.1$, $p = .00$, and a stricter alpha of .01 was utilized. An ANOVA for ELL type by district was conducted using 9th grade reading scores. Tables 25 summarizes results related to this analysis, and shows that there were significant differences between the ELL groups, $F(2, 15394) = 352.80$, $p < .01$ and district of attendance, $F(9, 15394) = 8.62$, $p < .01$.

Table 25

9th Grade Reading Scores by ELL Type and District (N = 15,424)

Source	*df*	*F*	Sum of Squares	Mean Square
ELL Type	2	352.80*	11664325.58	5832162.79
District	9	8.62*	1281773.28	142419.25
ELL Type x District	18	5.14*	1528940.94	84941.16
Error	15394		254479976.32	(16531.11)

Note: *$p < .01$

A post-hoc Tukey's HSD analysis was conducted to examine the relative performance of each ELL group. Similar to 7th grade reading scores, results show that non-ELLs performed the highest (M = 2291.78) followed by former-ELLs (M = 2279.01)

and long-term ELLs (M = 2181.73). Again, each group's means were significantly different from the others at $p < .05$. Table 26 illustrates the differences in 9th grade math score means between the ELL types.

Table 26

Differences in Means for 9th Grade Reading Scores by ELL Type

Contrast	Estimate	*SE*	95% Confidence Interval
Former ELL vs. Long ELL	97.28*	3.03	90.19, 104.37
Former ELL vs. Non-ELL	-13.77*	2.30	-19.16, -8.37
Non-ELL vs. Long-ELL	111.05*	3.21	103.53, 118.56

Note: *Significant at the .05 level using the Tukey HSD procedure.

9th grade math. Levene's test for the equality of variance results again indicated unequal population cell variances at the .05 level, $F(29, 15369) = 6.1$, $p = .00$, and a stricter alpha of .01 was utilized. An ANOVA for ELL type by district was conducted using 9th grade math scores. Tables 27 summarizes results related to this analysis, and shows that there were significant differences between the ELL types, $F(2, 15360) = 171.45$, $p < .01$ and district of attendance, $F(9, 15360) = 20.79$, $p < .01$.

Table 27

9th Grade Math Scores by ELL Type and District (N = 15,390)

Source	*df*	*F*	Sum of Squares	Mean Square
ELL Type	2	171.45*	15156098.68	7578049.34
District	9	20.79*	8268491.76	918721.31
ELL Type x District	18	3.57*	2840474.72	157804.15
Error	15360		678928079.56	(44201.05)

Note: *$p < .01$

A post-hoc Tukey's HSD analysis was conducted to examine the relative performance of each ELL group. Reflecting the pattern of 7th grade math scores, results show that former-ELLs performed the highest ($M = 2249.29$) followed by non-ELLs (M

= 2221.04) and long-term ELLs (M = 2132.02). Again, each group's means were significantly different from the others at $p < .05$. Table 28 illustrates the differences in 9th grade math score means between the ELL types.

Table 28

Differences in Means for 9th Grade Math Scores by ELL Type

Contrast	Estimate	*SE*	95% Confidence Interval
Former ELL vs. Long ELL	117.26*	4.94	105.68, 128.85
Former ELL vs. Non-ELL	28.25*	3.77	19.42, 37.08
Non-ELL vs. Long-ELL	89.02*	5.24	76.74, 101.30

Note: *Significant at the .05 level using the Tukey HSD procedure.

Summary. There exist significant group mean differences for ELL type on achievement scores in both reading and math, in 7th and 9th grades. However, the outcome ranks of ELL types differ depending on subject, with former-ELLs performing best in math, while non-ELLs perform best in reading in both years. Long-term ELLs performed poorest out of the three groups at both time points and subjects. The hypothesized outcome ranks outlined in Table 4 matched the results for reading at both time points; therefore, hypothesis 2 was partially confirmed. Table 29 illustrates the actual outcome ranks of the ELL types for both math and reading.

Table 29

Research Question 2 Results

Actual Reading Outcome Rank	ELL Type	Actual Math Outcome Rank	ELL Type
1.	Non-ELL	1.	Former-ELL
2.	Former-ELL	2.	Non-ELL
3.	Long-Term ELL	3.	Long-Term ELL

District Differences

There are clear performance differences between the 10 major urban school districts included in this study. The district factor was significant in all except one (9th grade reading by language programming and district) out of eight ANOVAs in the first two research questions. Furthermore, interactions between district and language programming or ELL type were significant in every analysis. Since district of enrollment was not initially part of this study's conceptualization, there are no hypotheses accompanying district performance. Table 30 presents the post-hoc analyses' homogenous subsets for each of the four study points, highlighting differential district performance from lowest to highest. Each subset is statistically significant from the others at $p < .05$. Figure 4 illustrates individual district performance at each study point.

Table 30

Homogenous Subsets for Mean Scores by District

7th Grade Reading		Subset				
District	*N*	1	2	3	4	5
2	1009	2193.78				
4	1713		2218.79			
5	1592		2233.04	2233.04		
1	627		2237.43	2237.43	2237.43	
6	3520		2239.57	2239.57	2239.57	
9	1543			2241.60	2241.60	
3	3371			2245.54	2245.54	
10	1336				2256.61	2256.61
8	970					2266.96
7	487					2271.61
7th Grade Math		Subset				
District	*N*	1	2	3	4	5
2	1003	2157.43				
5	1591	2171.91				
9	1545	2175.04	2175.04			
1	628		2193.79	2193.79		
6	3523			2205.99	2205.99	

Table 30 continued

4	1716			2208.20	2208.20	
3	3371			2211.37	2211.37	2211.37
8	970				2227.19	2227.19
10	1337					2232.28
7	487					2232.63

9th Grade Reading		Subset				
District	*N*	1	2	3	4	5
2	950	2245.79				
1	601	2259.73	2259.73			
4	1638	2259.80	2259.80			
5	1533	2260.12	2260.12			
3	3215		2265.56			
6	3358		2269.35	2269.35		
10	1306		2270.72	2270.72	2270.72	
7	451			2286.79	2286.79	2286.79
9	1438				2288.60	2288.60
8	934					2296.21

9th Grade Math		Subset				
District	*N*	1	2	3	4	5
1	595	2167.28				
2	959	2176.24	2176.24			
9	1426	2178.85	2178.85			
5	1532		2203.75	2203.75		
7	453			2212.01	2212.01	
8	932			2230.41	2230.41	2230.41
3	3204			2232.66	2232.66	2232.66
6	3354				2240.89	2240.89
10	1299					2242.77
4	1636					2245.16

Note: Subsets are significant at the .05 level using the Tukey HSD procedure. Group sizes are unequal, Type 1 error levels are not guaranteed.

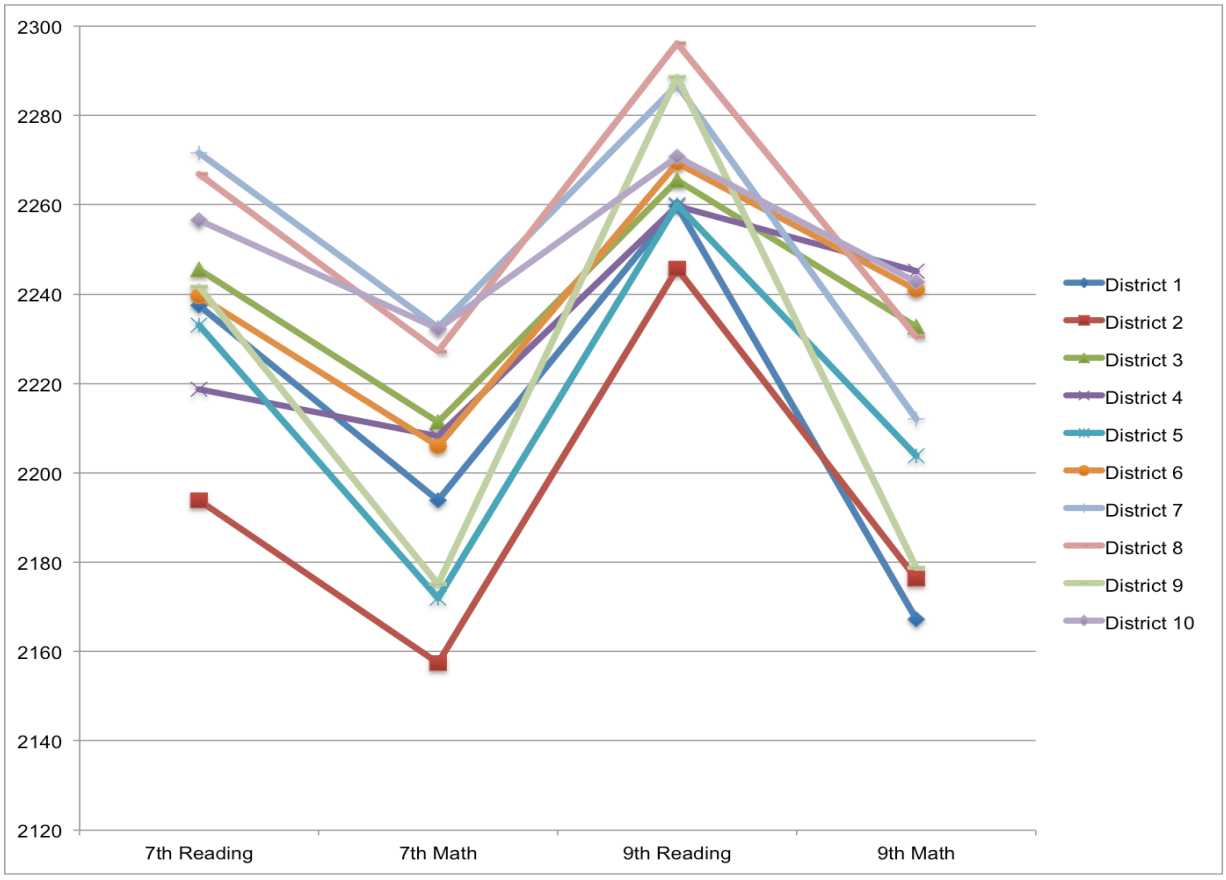

Figure 5. Mean achievement scores of each district at all study points.

Analyses and Results for Research Question 3

This is an exploratory analysis examining if significant mean differences in achievement between the two ELL groups exist across several independent variables at both time points. Hence, there are no directional hypotheses associated with this research question. Rather, it is an open-ended examination of various factors between the groups including: primary classroom model received, district of enrollment, and immigrant status.

This research question was examined in two parts: first (part A), using ANOVA to compare the two groups of ELL students to their non-ELL counterpart while including

language programming and immigrant status in the analysis. Second (part B), using ANCOVA to examine only the target ELL groups with the same additional independent variables in the analysis, and controlling for baseline English language proficiency by including TELPAS score as a covariate variable. Since the results for district and language program were already reported in the first two research questions, they are not reported again here. This research question focuses on differential achievement between former and long-term ELLs. Analyses focused on examining the group differences for immigrant status and possible interactions between ELL type and other independent variables. The adjusted means of ELL type with the addition of the covariate variable are also reported. Only two-way interactions were included in each analysis, as the research question does not address three-way interactions. Post-hoc one-way ANOVAs and independent samples t-tests were conducted to examine significant interactions. As with previous research questions, significant interactions including district were not further explored.

7th grade reading, part A. Prior to conducting the formal analysis of the data, Levene's test for the equality of variance was conducted. Results indicate that population cell variances are unequal at the .05 level, $F(120, 16047) = 1.8$, $p = .00$. A stricter alpha of .01 was utilized to control for the possible inflation of the Type 1 error rate. An ANOVA including the variables of district, ELL type, language program, and immigrant status was conducted using 7th grade reading scores. Table 31 summarizes the results related to this analysis, and shows that immigrant status was significantly associated with mean differences in achievement scores, $F(1, 16129) = 8.49$, $p < .01$. Immigrant students

(M = 2224.61; N = 1016) performed significantly higher than non-immigrant students (M = 2196.5; N = 15152). There was a significant interaction between ELL type and language programming, $F(3, 16129) = 4.59$, $p < .01$, but not between ELL type and immigrant status.

Table 31

7th Grade Reading Scores by District, ELL Type, Immigrant Status, and Language Program (N = 16,168)

Source	*df*	*F*	Sum of Squares	Mean Square
District	9	23.41*	4809194.63	534354.96
ELL Type	1	192.67*	4397105.10	4397105.10
Immigrant	1	8.49*	193639.63	193639.63
Language Program	3	5.94*	406705.92	135568.64
District x ELL Type	9	6.42*	2638378.80	146576.60
ELL Type x Immigrant	1	.79	36057.07	18028.53
ELL Type x Lang. Prog.	3	4.59*	314305.98	104768.66
Error	16129		368105775.03	(22822.60)

Note: *$p < .01$

Post-hoc analyses were conducted to examine the significant interaction between ELL type and language programming. Two separate one-way ANOVAs including language programming were conducted, with one examining former- ELLs and the other looking at long-term ELLs. For former-ELLs, there are significant group mean differences associated with language programming, $F(3, 8276) = 13.94$, $p < .01$. For long-term ELLs, $F(3, 2501) = 3.75$, $p = .011$, results barely miss the alpha cutoff for significance, yet language programming differences for this group were still examined as results were close to significant. Table 32 summarizes the sample size, mean scores, and standard deviations for the two types of ELL students who received each language

program model. Non-ELL students were not included in these post-hoc analyses; their means are reported below for reference.

Table 32

7th Grade Reading Means for ELL Type by Language Program

Lang. Prog.	ELL Type	
No Services	Former ELL	Long-Term ELL
N	724	43
M	2270.01	2182.49
SD	168.97	189.72
ESL		
N	2384	445
M	2257.36	2115.18
SD	156.07	147.38
Bilingual		
N	4257	1823
M	2259.93	2114.91
SD	142.34	149.10
Equal ESL/Bil.		
N	915	194
M	2227.90	2098.30
SD	149.665	141.46
No Services	Non-ELL	
N	5383	
M	2270.54	
SD	158.65	

Additional post-hoc comparisons using Tukey's HSD were conducted to examine group mean differences associated with language programming for each of the ELL groups. Results indicate that for former-ELLs, those who received equal ESL/bilingual education (M = 2227.90) had significantly lower scores than those who primarily received ESL (M = 2257.36), bilingual (M = 2259.93), or no services (M = 2270.01). For long-term ELLs, those who received equal ESL/bilingual (M = 2098.30), bilingual (M = 2114.91), and ESL (M = 2115.18) all had significantly lower scores than students who

received no services (M = 2182.49). Table 33 summarizes the mean differences for the two groups of ELLs by language programming. Figure 6 graphically presents the mean scores for each ELL type and language program.

Table 33

Differences in Means for 7th Grade Reading Scores by Language Programming for Former ELLs and Long-Term ELLs

Contrast	Estimate	*SE*	95% Confidence Interval
	Former ELLs		
Bilingual vs. ESL	2.58	3.82	-7.24, 12.39
Bilingual vs. No Svcs.	-10.08	6.00	-25.50, 5.35
Bilingual vs. Equal ESL/Bil.	32.04*	5.441	18.05, 46.02
ESL vs. No Svcs.	-12.65	6.336	-28.93, 3.63
ESL vs. Equal ESL/Bil.	29.46*	5.807	14.54, 44.38
No Svcs. vs. Equal ESL/Bil.	42.11*	7.427	23.03, 61.20
	Long-Term ELLs		
Bilingual vs. ESL	-.26	7.88	-20.52, 19.99
Bilingual vs. No Svcs.	-67.58*	22.99	-126.67, -8.48
Bilingual vs. Equal ESL/Bil.	16.61	11.25	-12.32, 45.53
ESL vs. No Svcs.	-67.31*	23.80	-128.45, -6.14
ESL vs. Equal ESL/Bil.	16.87	12.819	-16.08, 49.83
No Svcs. vs. Equal ESL/Bil.	84.18*	25.114	19.62, 148.75

Note: *Significant at the .05 level using the Tukey HSD procedure.

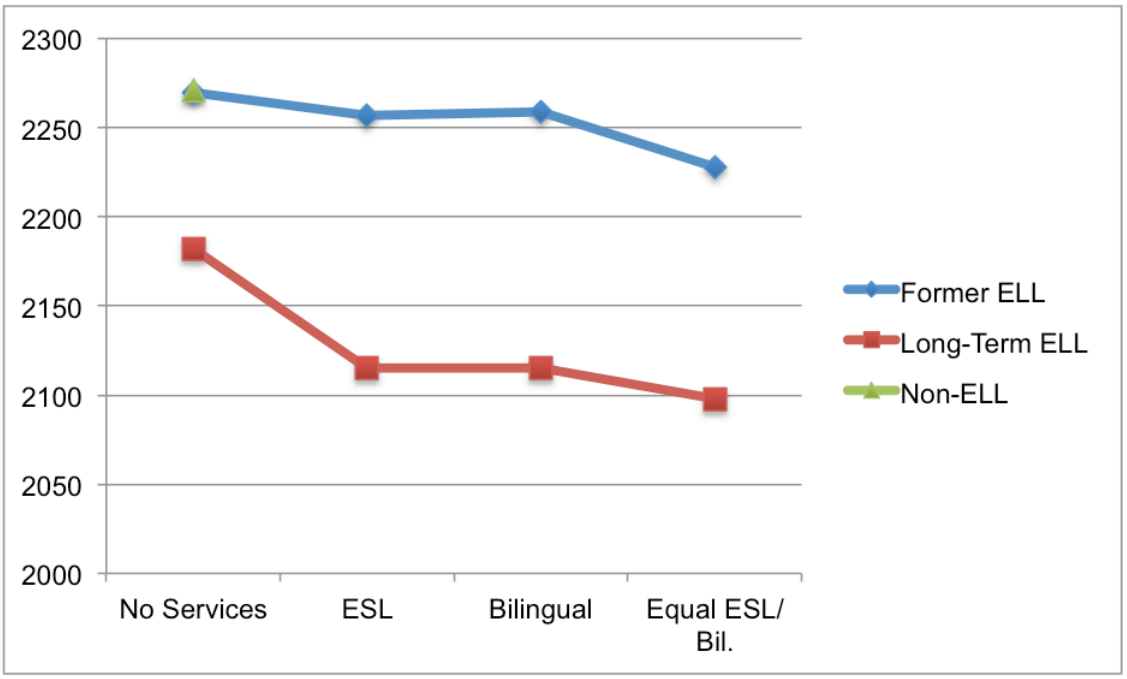

Figure 6. 7th grade reading means for each ELL type by language programming.

7th grade reading, part B. The results of Levene's test for the equality of variance, $F(102, 9333) = 1.16$, $p = .125$, indicated that the variance of the residuals was similar in each group, therefore a standard alpha level of .05 was utilized. An ANCOVA including a covariate variable of baseline TELPAS score and the independent variables of district, ELL type, language program, and immigrant status was conducted using 7th grade reading scores. Only ELL students were included in all ANCOVA analyses. Table 34 summarizes the results related to this analysis, and shows that the TELPAS score as a covariate was significant, $F(1, 9407) = 496.83$, $p < .05$, as were ELL type, $F(1, 9407) = 58.40$, $p < .05$ and immigrant status, $F(1, 9407) = 100.60$, $p < .05$. Taking into account baseline English proficiency, the adjusted mean for immigrants ($M = 2235.51$; 95% CI = 2221.97, 2249.04) was significantly higher than non-immigrants ($M = 2182.89$; 95% CI = 2173.24, 2192.55), as was the adjusted mean of former ELLs ($M = 2250.90$; 95% CI =

2242.08, 2259.71) compared to long-term ELLs (M = 2167.50; 95% CI = 2148.14, 2186.86). The interaction between ELL type and language program is also significant, $F(3, 9407) = 2.66, p < .05$.

Table 34

7th Grade Reading Scores by District, ELL Type, Immigrant Status, and Language Program with TELPAS as Covariate (N = 10,785)

Source	*df*	*F*	Sum of Squares	Mean Square
TELPAS	1	496.83*	9952708.29	9952708.29
District	9	22.37*	4032759.19	448084.35
ELL Type	1	58.40*	1169886.25	1169886.25
Language Program	3	10.22*	614226.31	204742.10
Immigrant	1	100.60*	2015197.58	2015197.58
District x ELL Type	9	4.47*	805548.39	89505.38
ELL Type x Lang. Prog.	3	2.66*	159642.19	53214.06
ELL Type x Immigrant	1	3.212	64493.44	64493.44
Error	9407		188446264.80	(20032.56)

Note: *$p < .05$.

Post-hoc analyses were conducted to examine the significant interaction between ELL type and language programming while taking into account the effects of the TELPAS covariate. A general linear model with covariate was utilized in order to examine all pairwise comparisons in one analysis. Bonferroni post-hoc test of adjusted means compared the estimated marginal means for language programming groups at each level of ELL type. Table 35 displays the adjusted mean differences, standard errors, and 95% confidence intervals for these comparisons. Figure 7 displays the adjusted means for each language program group by ELL type.

Table 35

Differences in Adjusted Means for 7th Grade Reading by Language Program and ELL Type with TELPAS Covariate

Contrast	Estimate	*SE*	95% Confidence Interval
	Former ELLs		
Bilingual vs. ESL	46.38*	4.26	38.03, 54.87
Bilingual vs. No Svcs.	39.81*	7.69	24.74, 54.87
Bilingual vs. Equal ESL/Bil.	51.32*	5.50	40.53, 62.11
ESL vs. No Svcs.	-6.57	7.95	-22.16, 9.02
ESL vs. Equal ESL/Bil.	4.95	5.98	-6.77, 16.67
No Svcs. vs. Equal ESL/Bil.	11.52	8.72	-5.58, 28.62
	Long-Term ELLs		
Bilingual vs. ESL	19.48*	7.846	4.10, 34.86
Bilingual vs. No Svcs.	-41.67	28.51	-97.55, 14.21
Bilingual vs. Equal ESL/Bil.	37.38*	11.25	15.33, 59.44
ESL vs. No Svcs.	-61.15*	29.09	-118.18, -4.12
ESL vs. Equal ESL/Bil.	17.90	12.78	-7.14, 42.95
No Svcs. vs. Equal ESL/Bil.	79.05*	30.19	19.87, 138.23

Note: *Significant at the .05 level using the Bonferroni procedure. Covariate value in the model is evaluated at TELPAS = 2.78

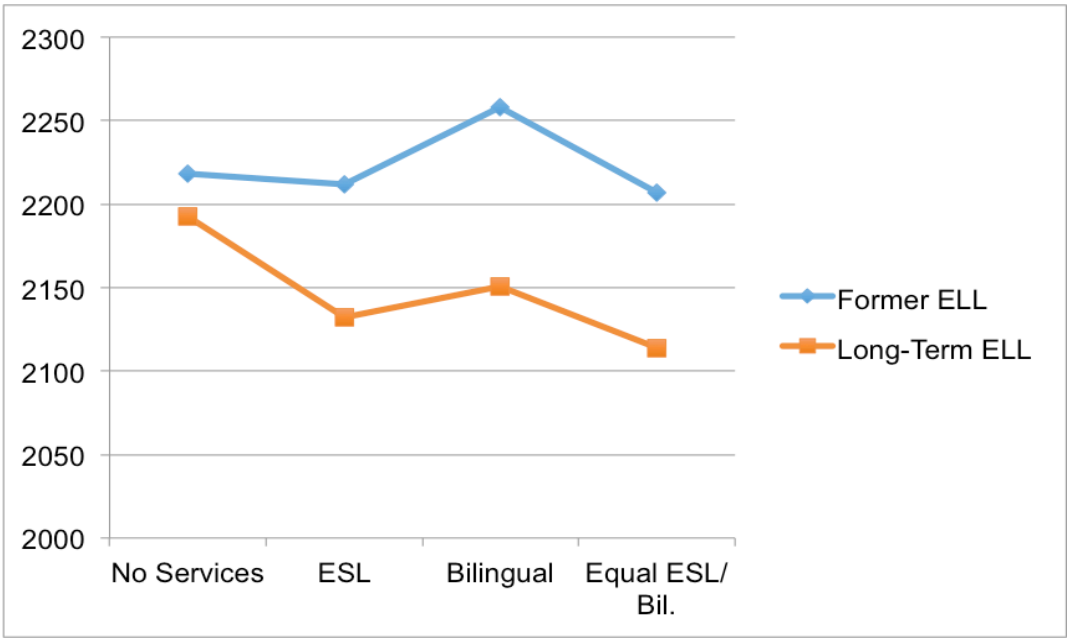

Figure 7. 7th grade reading adjusted means for ELL type by language programming with TELPAS covariate.

Results for former-ELLs show that there were no significant differences in adjusted means between ESL and no services; ESL and equal ESL/bilingual; no services and equal ESL/bilingual. However, former-ELLs in bilingual classes performed significantly higher ($p < .05$) than all three of the other language program models after adjusting for baseline TELPAS score.

For long-term ELLs, results show no significant differences in adjusted means between ESL vs. equal ESL/bilingual, and bilingual vs. no services. Students who primarily received bilingual education scored significantly higher ($p < .05$) than students in ESL, and equal ESL/bilingual. Finally, students who did not receive any services scored significantly higher ($p < .05$) than students who received ESL and students who received equal ESL/bilingual.

7th grade math, part A. Levene's test for the equality of variance indicated that population cell variances are unequal at the .05 level, $F(120, 16050) = 2.0$, $p = .00$, and an alpha of .01 was utilized. An ANOVA including the variables of district, ELL type, language program, and immigrant status was conducted using 7th grade math scores. Table 36 summarizes the results related to this analysis. There were no significant group mean differences related to immigrant status, nor was there a significant interaction between ELL type and immigrant status. The interaction between ELL type and language programming was almost significant, $F(3, 16132) = 3.69$, $p = .011$. Therefore, post-hoc analyses examining this interaction were still conducted, as the result was close to significance.

Table 36

7th Grade Math Scores by District, ELL Type, Immigrant Status, and Language Program (N = 16,171)

Source	*df*	*F*	Sum of Squares	Mean Square
District	9	32.74*	7250573.15	805619.24
ELL Type	1	92.32*	2271857.77	2271857.77
Immigrant	1	1.88	46144.98	46144.98
Language Program	3	6.68*	493210.35	164403.45
District x ELL Type	9	4.36*	1932364.37	107353.58
ELL Type x Immigrant	1	.487	23950.26	11975.13
ELL Type x Lang. Prog.	3	3.69[a]	272359.30	90786.43
Error	16132		396973222.45	(24607.81)

Note: *$p < .01$, [a]$p = .011$.

Post-hoc analyses were conducted to examine the interaction between ELL type and language programming in 7th grade math scores. Two separate one-way ANOVAs including language programming were conducted, with one examining former- ELLs and the other looking at long-term ELLs. The association between language programming and achievement is significant for both former-ELLs, $F(3, 8278) = 19.70$, $p < .01$, and long-term ELLs, $F(3, 2503) = 10.50$, $p < .01$. Table 37 summarizes the sample size, mean scores, and standard deviations for the two types of ELL students who received each language program model in comparison to their non-ELL peers.

Table 37

7th Grade Math Means for ELL Type by Language Program

Lang. Prog.	ELL Type	
No Services	Former ELL	Long-Term ELL
N	719	43
M	2226.40	2162.28
SD	171.60	166.40
ESL		
N	2384	444
M	2210.08	2097.86
SD	165.57	136.82
Bilingual		
N	4260	1824
M	2232.51	2132.32
SD	159.96	147.96
Equal ESL/Bil.		
N	916	193
M	2193.56	2094.49
SD	158.10	126.30
No Services	Non-ELL	
N	5388	
M	2208.21	
SD	159.54	

Additional post-hoc comparisons using Tukey's HSD were conducted to examine the mean differences of language program models within each of the ELL groups. Results indicate that for former-ELLs, those who primarily received bilingual education (M = 2232.51) scored significantly higher than ESL (M = 2210.08) and equal ESL/bilingual (M = 2193.56) at $p < .05$. Students who primarily received ESL scored significantly higher than those in equal ESL/bilingual. Former-ELLS in the equal ESL/bilingual group scored significantly lower than all other groups. Students who received no services (M = 2226.40) did not score significantly differently than students primarily in ESL or bilingual. For long-term ELLs, students who received no services (M = 2162.28) scored

significantly higher than students who primarily received ESL (M = 2097.86) or equal ESL/bilingual (M = 2094.49) at $p < .05$. Long-term ELLs who primarily received bilingual (M = 2132.32) scored significantly higher than ESL and equal ESL/bilingual. There were no statistical differences between bilingual and no services, nor between ESL and equal ESL/bilingual. Table 38 summarizes the mean differences for each group of ELLs by language programming. Figure 8 graphically presents the mean scores for each ELL type and language program.

Table 38

Differences in Means for 7th Grade Math Scores by Language Programming for Former ELLs and Long-Term ELLs

Contrast	Estimate	*SE*	95% Confidence Interval
	Former ELLs		
Bilingual vs. ESL	22.43*	4.15	11.76, 33.11
Bilingual vs. No Svcs.	6.11	6.55	-10.71, 22.94
Bilingual vs. Equal ESL/Bil.	38.96*	5.92	23.76, 54.16
ESL vs. No Svcs.	-16.32	6.91	-34.08, 1.44
ESL vs. Equal ESL/Bil.	16.52*	6.31	.30, 32.75
No Svcs. vs. Equal ESL/Bil.	32.84*	8.09	12.05, 53.64
	Long-Term ELLs		
Bilingual vs. ESL	34.46*	7.66	14.76, 54.16
Bilingual vs. No Svcs.	-29.96	22.34	-87.40, 27.48
Bilingual vs. Equal ESL/Bil.	37.83*	10.96	9.65, 66.01
ESL vs. No Svcs.	-64.42*	23.13	-123.88, -4.96
ESL vs. Equal ESL/Bil.	3.37	12.49	-28.72, 35.47
No Svcs. vs. Equal ESL/Bil.	67.79*	24.42	5.01, 130.57

Note: *Significant at the .05 level using the Tukey HSD procedure.

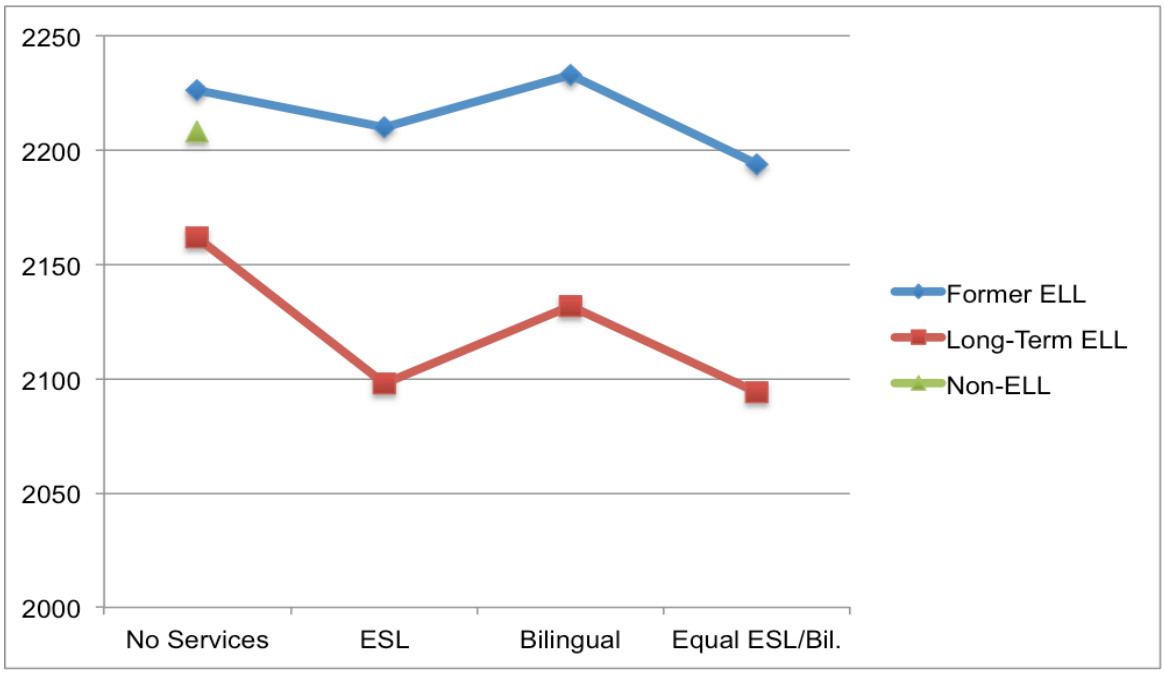

Figure 8. 7th grade math means for each ELL type by language programming.

7th grade math, part B. Levene's test for the equality of variance indicated that population cell variances are unequal at the .05 level, $F(102, 9335) = 1.75$, $p = .00$, and an alpha of .01 was utilized. An ANCOVA including a covariate variable of baseline TELPAS score and the independent variables of district, ELL type, language program, and immigrant status was conducted using 7th grade math scores. Table 39 summarizes the results related to this analysis, and shows that TELPAS as a covariate was significant, $F(1, 9409) = 224.61$, $p < .01$, as were ELL type, $F(1, 9409) = 27.01$, $p < .01$ and immigrant status, $F(1, 9409) = 70.07$, $p < .01$. Taking into account baseline English proficiency, the adjusted mean for former ELLs ($M = 2219.40$; 95% CI = 2209.91, 2228.88) was significantly higher than long-term ELLs ($M = 2158.29$; 95% CI = 2137.43, 2179.15), and adjusted mean for immigrants ($M = 2212.46$; 95% CI = 2197.89, 2227.02)

was significantly higher than non-immigrants (M = 2165.23; 95% CI = 2154.82, 2175.64). There were no significant interactions between independent variables.

Table 39

7th Grade Math Scores by District, ELL Type, Immigrant Status, and Language Program with TELPAS as Covariate (N = 10,783)

Source	*df*	*F*	Sum of Squares	Mean Square
TELPAS	1	224.61*	5205516.82	5205516.82
District	9	27.30*	5693754.59	632639.40
ELL Type	1	27.02*	626140.17	626140.17
Language Program	3	11.50*	799350.05	266450.02
Immigrant	1	70.07*	1623856.59	1623856.59
District x ELL Type	9	1.57	326850.23	36316.69
ELL Type x Lang. Prog.	3	2.06	143340.15	47780.05
ELL Type x Immigrant	1	1.22	28191.42	28191.42
Error	9409		218061519.67	(23175.844)

Note: $*p < .01$.

9th grade reading, part A. Levene's test for the equality of variance indicated that population cell variances are unequal at the .05 level, $F(119, 15304) = 1.34$, $p = .01$, and an alpha of .01 was utilized. An ANOVA including the variables of district, ELL type, language program, and immigrant status was conducted using 9th grade reading scores. Table 40 summarizes the results related to this analysis. There were no significant group differences for immigrant status, nor were there significant interactions between ELL type and language programming or between ELL type and immigrant status.

Table 40

9th Grade Reading Scores by District, ELL Type, Immigrant Status, and Language Program (N = 15,424)

Source	*df*	*F*	Sum of Squares	Mean Square
District	9	7.82*	1161219.83	129024.43
ELL Type	1	134.77*	2223597.10	2223597.10
Immigrant	1	1.50	24796.35	24796.35
Language Program	3	2.16	106708.51	35569.50
District x ELL Type	9	4.69*	1393237.51	77402.08
ELL Type x Immigrant	1	.03	1104.57	552.28
ELL Type x Lang. Prog.	3	.32	15691.80	5230.60
Error	15385		253845909.57	(16499.57)

Note: $*p < .01$.

9th grade reading, part B. Levene's test for the equality of variance indicated that population cell variances are unequal at the .05 level, $F(102, 8954) = 1.33$, $p = .01$, and an alpha of .01 was utilized. An ANCOVA including a covariate variable of baseline TELPAS score and the independent variables of district, ELL type, language program, and immigrant status was conducted using 9th grade reading scores. Table 41 summarizes the results related to this analysis, and shows that the TELPAS score as a covariate was significant, $F(1, 9028) = 267.18$, $p < .01$, as were ELL type, $F(1, 9028) = 48.76$, $p < .01$ and immigrant status, $F(1, 9028) = 54.82$, $p < .01$. Taking into account baseline English proficiency, the adjusted mean for former ELLs ($M = 2274.86$; 95% CI = 2267.05, 2282.67) was significantly higher than long-term ELLs ($M = 2207.18$; 95% CI = 2189.98, 2224.37), and the adjusted mean for immigrants ($M = 2258.34$; 95% CI = 2246.30, 2270.37) was significantly higher than non-immigrants ($M = 2223.70$; 95% CI = 2215.13, 2232.27). There were no significant interactions of interest.

Table 41

9th Grade Reading Scores by District, ELL Type, Immigrant Status, and Language Program with TELPAS as Covariate (N = 10,276)

Source	*df*	*F*	Sum of Squares	Mean Square
TELPAS	1	267.18*	4084632.21	4084632.21
District	9	8.68*	1193883.76	132653.75
ELL Type	1	48.76*	745426.04	745426.04
Language Program	3	5.09*	233408.30	77802.77
Immigrant	1	54.82*	838108.38	838108.38
District x ELL Type	9	4.22*	580015.44	64446.16
ELL Type x Lang. Prog.	3	.39	17647.39	5882.46
ELL Type x Immigrant	1	.68	10323.67	10323.67
Error	9028		138021007.58	(15288.11)

Note: $*p < .01$.

9th grade math, part A. Levene's test for the equality of variance indicated that population cell variances are unequal at the .05 level, $F(119, 15270) = 1.97$, $p = .00$, and an alpha of .01 was utilized. An ANOVA including the variables of district, ELL type, language program, and immigrant status was conducted using 9th grade math scores. Table 42 summarizes the significant results related to this analysis. There were significant group differences for immigrant status, $F(1, 15351) = 8.86$, $p < .01$. Immigrant students ($M = 2214.41$; $N = 979$) scored significantly higher than non-immigrants ($M = 2177.61$; $N = 14411$). There were no significant interactions between ELL type and language programming or between ELL type and immigrant status.

Table 42

9th Grade Math Scores by District, ELL Type, Immigrant Status, and Language Program (N = 15,390)

Source	*df*	*F*	Sum of Squares	Mean Square
District	9	19.51*	7742093.83	860232.65
ELL Type	1	67.49*	2975481.53	2975481.53
Immigrant	1	8.86*	390608.71	390608.71
Language Program	3	3.27	432775.55	144258.52
District x ELL Type	9	2.67*	2121688.28	117871.57
ELL Type x Immigrant	1	1.68	147650.94	73825.47
ELL Type x Lang. Prog.	3	1.00	132155.38	44051.79
Error	15351		676766558.86	(44086.16)

Note: $*p < .01$

9th grade math, part B. Levene's test for the equality of variance indicated that population cell variances are unequal at the .05 level, $F(101, 8938) = 1.72$, $p = .00$, and an alpha of .01 was utilized. An ANCOVA including a covariate variable of baseline TELPAS score and the independent variables of district, ELL type, language program, and immigrant status was conducted using 9th grade math scores. Table 43 summarizes the results related to this analysis, and shows that TELPAS as a covariate was significant, $F(1, 9011) = 170.17$, $p < .01$, as were ELL type, $F(1, 9011) = 22.83$, $p < .01$ and immigrant status, $F(1, 9011) = 53.14$, $p < .01$. Taking into account baseline English proficiency, the adjusted mean of former ELLs ($M = 2233.16$; 95% CI = 2220.27, 2246.05) was significantly higher than long-term ELLs ($M = 2154.76$; 95% CI = 2125.50, 2184.03), as was the adjusted mean of immigrants ($M = 2221.99$; 95% CI = 2201.82, 2242.16) compared with non-immigrants ($M = 2165.93$; 95% CI = 2151.36, 2180.50). There were no significant interactions between independent variables.

Table 43

9th Grade Math Scores by District, ELL Type, Immigrant Status, and Language Program with TELPAS as Covariate (N = 10,254)

Source	*df*	*F*	Sum of Squares	Mean Square
TELPAS	1	170.17*	7041635.19	7041635.19
District	9	16.92*	6300387.55	700043.06
ELL Type	1	22.83*	944848.45	944848.45
Language Program	3	7.36*	913492.38	304497.46
Immigrant	1	53.14*	2199009.41	2199009.41
District x ELL Type	9	.72	268871.15	29874.57
ELL Type x Lang. Prog.	3	.45	55294.53	18431.51
ELL Type x Immigrant	1	2.50	103509.07	103509.07
Error	9011		372886761.51	(41381.29)

Note: $*p < .01$

Summary. When all students were included in the analysis (part A), there were significant interactions between ELL type and language programming for 7th grade reading and close to significant interactions for 7th grade math scores. The performance of both former and long-term ELLs was moderated by the primary language program model students received in elementary school. By the time students reached 9th grade, there were no longer significant interactions between ELL type and language programming. Immigrant status was significant for 7th grade reading and 9th grade math scores, with immigrants performing higher than non-immigrants.

When examining 7th grade reading scores for all students, results indicate that former-ELLs who received equal ESL/bilingual services performed significantly lower than all other program models. For long-term ELLs, results were close to significant. Students who received no services performed higher than all other program models. As for 7th grade math scores, the result for group differences was close to significant. Former-ELLs who received equal ESL/bilingual services again performed lower than all

other program models. However, former-ELLs who primarily received bilingual services performed higher than those in primary ESL or equal ESL/bilingual. Former-ELLs who received no services did not perform differently than those primarily in ESL or bilingual. Long-term ELLs who received no services or primary bilingual education performed higher than those who received primary ESL or equal ESL/bilingual.

When the analyses included only ELL students and baseline TELPAS score was added as a covariate variable (part B), results show that there exist significant group mean differences for ELL type and immigrant status at every point of the study. Controlling for baseline English language proficiency, former-ELLs performed higher than long-term ELLs, and immigrant students performed higher than non-immigrants at all points. The interaction between ELL type and language programming was only significant for 7th grade reading. For former-ELLs at this time point, those who primarily received bilingual education performed significantly higher than all other language program models. Long-term ELLs who primarily received bilingual services performed significantly better than those in ESL and equal ESL/bilingual. However, there was no significant difference between bilingual long-term ELLs and those who did not receive any services. The no service group performed significantly higher than both ESL students and those who received equal ESL/bilingual.

Chapter Five: Discussion

English-language-learners represent an ever-growing population of American students, educated in a variety of different language programming models ranging from no specialized language services to bilingual education. The existing literature on the achievement outcomes of these students primarily supports bilingual education as the most promising language-programming model for ELLs to gain equal footing with their native English-speaking peers across academic subjects (August & Shanahan, 2006; Genesee, Lindholm-Leary, Saunders, & Christian, 2006; Goldenberg, 2008; Mora, Wink, & Wink, 2001; Thomas & Collier, 1997; 2002). The theoretical framework of a common underlying proficiency that serves to connect linguistic knowledge between similar first and second language skills, such as the case between Spanish and English, lends further support for the superiority of bilingual education programs (Cummins, 1981). Prior research has failed to longitudinally study a demographically similar cohort of ELLs as they progress through the educational system. This study aims to address that gap.

The primary focus of the current study was to assess the relative effectiveness of various models of elementary language programming as measured by student achievement outcomes during secondary years. Unexpected results were found, first, with ELLs who did not receive any language services performing comparably to their non-ELL peers, and significantly higher than students who received language services. Secondly, out of ELLs who did receive language services, those who were primarily enrolled in ESL services performed higher students who primarily received bilingual education or equal years of ESL and bilingual programming. The effects of language

program model received were significant for secondary reading scores, but not for math scores. Furthermore, as students progressed in their education, the achievement differences between students who primarily received ESL and bilingual models became insignificant.

A secondary purpose of this study was to explore the factors leading to differential achievement outcomes between former-ELLs who are able to exit language programming after gaining proficiency in English during elementary years, and long-term ELLs who remain in ELL status through secondary schooling. As the literature predicted, former-ELLs performed significantly higher than long-term ELLs at both time points in reading and math. When compared to non-ELLs, results found that non-ELLs perform better in reading, but former-ELLs perform better in math at both time points. After exploring the factors contributing to differential achievement between the two ELL types, results show that significant (or near significant) interactions exist between language model received and ELL type in 7th grade, but not 9th. Again, students who did not receive any services appear to perform just as well or better than students who received language programming, especially for long-term ELLs. After controlling for the effects of baseline English language proficiency, there remained significant differences between ELL types at all study points; however, interactions between language programming and ELL type only remained for 7th grade reading. Results for these scores lend support for bilingual programming for both ELL groups, yet, long-term ELLs who did not receive any services also performed comparably to those who received bilingual education and higher than the other two language groups.

Finally, this study included data from students who met sampling criteria enrolled in all 10 major urban Texas school districts during the years of study. Results found significant differences in ELL composition, language model provision, and achievement outcomes between the school districts. This was an unintended result of the current study; nevertheless, it provides important context for the findings and implications of this study.

Key Findings

No services group. Primarily, this study provided evidence that ELLs who did not receive any language services performed comparably to, or higher than, demographically similar non-ELL students. This finding starkly contradicts existing literature on predictors of ELL achievement. Thomas and Collier's (2002) landmark study on language programming models suggests that students who are mainstreamed into English-only programs have the poorest outcomes in reading and math achievement by 5th grade, and the highest number of dropouts. When considering current results in the context of existing literature, it is first important to dissect the composition of the students who did not receive services and explain how this variable was derived. The "no services" group was created from students who were categorized as ELL at any point during data availability (2nd through 9th grades), who did not enroll in language programming in any year of their elementary schooling. Out of the total study sample, only 6.6% of ELL students did not receive any language services. When examining their LEP permission data (Table 7), it appears that the majority of these students' parents denied special language programming at the start of their education. In view of these results, what are possible explanations for these students' relatively high achievement?

One hypothesis is that since these students began and spent the duration of their educations with native English speakers, they were provided more opportunities to socialize and practice English skills from a young age. It is also possible that the mainstream English classroom composition itself differs from language service classrooms in the quality of teacher performance, income-level of classroom peers, and curriculum opportunities afforded to learners.

It is tempting to attribute the differential performance of this group to an underlying student characteristic such as cognitive ability or perseverance. While these student-level differences may indeed play a role, they do not explain the significantly higher scores of long-term ELLs who did not receive any services. Long-term ELLs by definition are students who have trouble achieving adequate English proficiency within normal limits. It can be expected that student characteristics of achievement, such as ability, would apply to the timely acquisition of English language proficiency. Yet, when examining the differential outcomes of this group of underachievers, those who never received language services continue to perform better than students enrolled in language programming.

The exclusion of students who receive special education from the study may be another possible explanation for these unexpected results. Out of all students who met demographic sampling criteria, 15.1% were excluded due to enrollment in special education at any point during the study's timeframe. It is possible that struggling students who did not receive language services were instead placed in special education as a way to meet their language needs, thus leaving the remaining students who received no

services as the higher ability group. The overlap between special education and ELLs is further explored in the Limitations section below.

Bilingual education group. While the literature supports bilingual education as the most promising language program model for student achievement outcomes (August & Shanahan, 2006; Genesee, Lindholm-Leary, Saunders, & Christian, 2006; Goldenberg, 2008; Mora, Wink, & Wink, 2001; Thomas & Collier, 1997; 2002), the current study did not support these findings. In fact, with regard to both 7th and 9th grade reading scores, ELLs primarily enrolled in ESL during elementary years scored higher than students primarily enrolled in bilingual, although this difference was only statistically significant during 7th grade. Interestingly, although there were no significant group differences for math scores, students who received bilingual services performed higher in math than ESL students at both time points.

While bilingual education as a whole did not perform as highly as expected, it is important to note that when baseline English language proficiency was accounted for, former-ELLs enrolled in bilingual education performed significantly higher in 7th grade reading outcomes than all other language program models. However, this difference was not maintained as students progressed to high school. It is possible that bilingual education is most effective in the short-term for a specific subset of students: those who are able to exit language programming within the recommended timeframe, and with an expected baseline English proficiency. Perhaps bilingual programs spend less time on differentiation, and more effort meeting the needs of the "average" ELL.

In speculating about the relatively low performance of bilingual models, the composition of these classrooms was further explored through examining the non-ELL comparison group. The literature supports that students who are enrolled in two-way dual-language immersion classrooms comprised of both ELLs and native-English speakers generally have positive achievement outcomes (Krashen, 2004; Thomas & Collier, 2002). Since data delineating the specific subtypes of bilingual education programs were not available at the start of this study, it is impossible to calculate the ratios of students who received each type of bilingual education. However, out of 5875 non-ELL students, 5866 primarily received no language services during their education. This suggests that hardly any non-ELL students potentially participated in a two-way dual-language immersion bilingual program during the study's timeframe. It is also possible that the PEIMS indicator for participation in bilingual education was only used for ELL students at that time. However, if the former is true and only 11 non-ELL students participated in bilingual services, this finding highlights the stark contrast between research and practice. Lara-Alecio, Galloway, Irby, Rodriquez, and Gomez (2004) found that 10% of Texas school districts reported the implementation of dual-language programs in the early 2000's. The current findings show that in practice, this number may be even lower in Texas's major urban school districts. The homogeneity of bilingual programs compared to ESL or mainstream English programs might be another explanation for bilingual programming's lower achievement outcomes.

Other language program groups. Students who primarily received ESL perform significantly higher than students who primarily received bilingual on 7th grade reading

outcomes, but by and large, there were no significant differences between ESL and bilingual education programming on the remaining three study points. This result demonstrates that in practice, student achievement between the two main language programming models are comparable.

Results show that students who received equal ESL and bilingual programming during elementary school consistently performed the lowest at both time points and subjects. Considering the diversity of available language programming between districts, with some clearly preferring one model over others, it is expected that individual schools within these districts would generally offer one program model over others as well. Thus, for students to receive equal years of both ESL and bilingual services, it is likely that these are also students who demonstrate the greatest school mobility. Even if these students remained at the same campus, the lack of consistency within their education might contribute to lower achievement outcomes. Prior research demonstrate that students who received inconsistent schooling such as frequently changing schools or moving between language programming models, possibly due to a migrant lifestyle or poverty, are at greater risk for poor achievement outcomes (Menken & Kleyn, 2009; Calderon, Slavin & Sanchez, 2011).

Based on the literature, it was predicted that significant differences between language program models would exist for both math and reading achievement outcomes (Thomas & Collier, 2002). The current study found that group differences only exist for reading outcomes and not math. Considering that the linguistic complexity for reading assessments is expected to be more advanced than for math assessments, it is likewise

expected that higher cognitive academic language proficiency (CALP) skills are required to access the material. This may explain the presence of group differences for reading scores but not math scores. Perhaps the primary focus of ESL classrooms is on rapid English language acquisition, leading to better developed CALP skills and associated higher reading scores at an earlier age. This difference appears to even out over the course of students' education, leading to no significant differences between primary ESL and primary bilingual by the time students enter high school.

Secondary Findings

ELL groups. The current study confirmed findings from prior research that former-ELLs score higher than long-term ELLs on both math and reading assessments during secondary education (Menken & Kleyn, 2009; Olsen, 2010). Long-term ELLs consistently scored the lowest out of all students groups, and the achievement results of some long-term ELL subgroups are particularly disheartening. The passing score on the TAKS is 2100. Long-term ELLs who received equal years of ESL and bilingual programming had a mean score in the failing range for both math (M = 2094.49) and reading (M = 2098.30) during 7th grade. Considering the role that high-stakes assessment scores play in a host of outcomes such as school-wide accountability, grade retention, and federal funding (NCLB, 2002), the failing scores of these students have far-reaching implications.

On a more positive note, results show that former-ELLs are performing on par with their non-ELL peers and even outperforming them on measures of math achievement. Although there remain significant differences between non-ELL and

former-ELL achievement scores, the gap between the two groups was only around 20 points, compared to over 100 points for the mean group differences between non-ELL and long-term ELL students. Current results confirm prior research that suggest former-ELLs score higher than their non-ELL peers on math, but results do not support this finding for reading outcomes (Ardasheva, Tretter, & Kinny, 2012; Kim, Herman, & National Center for Research on Evaluation, 2009). It is possible that the same ability required for students to achieve English language proficiency in a relatively short amount of time also carries through to other areas of achievement such as math.

Immigrant status. This study found that immigrant students as a whole performed better than non-immigrants in 7th grade reading and 9th grade math outcomes. When controlling for the effects of baseline English proficiency, immigrant students scored higher than non-immigrants at all study points. Current results confirm previous research indicating that U.S. born students tend to perform lower on achievement measures than their immigrant counterparts (Slama, 2012). Interestingly, there were no significant interactions between ELL-type and immigrant status at any of the study points. Prior research has found that U.S. born students were more likely to become long-term ELLs than immigrant students (Slama, 2012), but this difference was not found in the current study.

District differences. An unintended finding of this study was the drastic differences between school districts in their composition of students, available language programming types, and subsequent achievement outcomes. The primary finding was that there exist considerable differences between school districts in the type of language

programming each offers. Most districts appeared to primarily provide bilingual programming (district 2, district 4, district 6, district 7, district 8, district 9, district 10), with less than 6% of students enrolled in ESL in each of these districts. However, the second largest urban district in Texas (district 3) provided mostly ESL programming, with over half of this district's students primarily enrolled in ESL. The remaining districts have students primarily enrolled in either ESL or bilingual in nearly equal proportions (district 1, district 5). With such diversity in local practice, it becomes difficult to make generalizations about a superior language model across Texas. Because most urban districts offer one primary language program model, the available programming options for students within each district becomes limited.

When examining the student composition of districts, similar patterns emerged. Half of the districts have former-ELLs as the dominant group (district 1, district 3, district 4, district 5, district 6), while three districts have non-ELLs comprising the majority of their student population (district 7, district 8, district 9). District 2 is the only district with relatively equal distribution of students between all ELL types. District 10 is comprised of 47% non-ELLs, and has relatively equal distribution of former and long-term ELLs. Another way that districts differed from each other is their apparent size. While district 6 was comprised of 3520 target students, district 7 only had 487 students who met sampling criteria. While it is likely that districts actually differ in size, it is also possible that districts with a smaller study sample are comprised of more students who did not meet sampling criteria, for example, students of different races and higher income levels.

The heterogeneity of student composition within districts may be another contributing factor to differential achievement outcomes.

It is important to keep district diversity in mind when examining the relative performance of each district. For example, districts 7 and 8 performed in the top quartile of districts at all study points except 9th grade math. However, they are comprised primarily of non-ELL students, who tend to perform higher than ELL students, especially in reading. On the other hand, district 9 is also comprised of mostly non-ELL students, and its relative performance varies from bottom to top across study points. Perhaps the proportion of long-term ELL students within a district can serve as an informal measure of district performance. Current findings confirm that long-term ELLs tend to perform lower on all achievement outcomes compared to their former and non-ELL counterparts. District 2 was the only district with equal division between ELL types. It had the highest relative proportion of long-term ELLs between districts (31%), and it also consistently performed near the bottom of all districts across study points. Notably, district 10 also had higher proportions of long-term ELLs than most districts (29%), but it consistently performed in the top half of districts across study points.

Specific district performance aside, it is important to consider other possible contributions to variable student achievement across the state. A recent study conducted by the Martin Prosperity Institute at the University of Toronto (2015) examined the overall economic segregation of over 350 metropolitan areas across the United States and Canada. Results found that Texas cities comprised four out of the top ten large metros with the highest segregation levels (Florida & Mellander, 2015). This study included

income, educational, and occupational segregation in the creation of the overall economic segregation measure. Implications suggest that many of the school districts included in the current study were located in areas of high economic segregation. Meaning that individuals with low-income, working class, high school educated or below backgrounds, tended to live, work, and go to school in areas separate from those who were more prosperous and educated. Since the current study only included low-income, Latino students in the sample, it is possible that many of these students lived in areas of high economic segregation, which may be partially responsible for their lower achievement compared to students who lived in areas with less economic segregation. Further exploration of this hypothesis would require identification of specific districts and metropolitan areas.

Ultimately, there exists significant variability in district composition, available language models, and relative achievement across the years of study, making it impossible to extrapolate conclusions about each district's relative performance from the current data. Furthermore, the significant interactions involving district of enrollment were not examined in this study, leaving many questions unanswered and the possibility open for future research.

Implications of Findings

The key implications of current findings focus on the relatively low performance of bilingual education compared to other language program models, as well as the surprisingly high achievement of students who received no specialized language services. This pattern highlights the need for additional, thorough, longitudinal research on the

long-term academic trajectories of students enrolled in various types of language programs. It is possible that prior research examined language programming in a contained and optimal environment afforded to short-term studies and/or experimental designs, whereas this is rarely the case in practice. The confounding factors of poverty, student mobility, and diverse local practices make it difficult to ensure that language program models were delivered with fidelity and consistency. Regardless, if bilingual education in practice is truly not producing the high levels of achievement that has been established in prior research, this opens the door for considerable policy change. It appears that policy shifts away from bilingual education in Texas are already in play, with the ending of developmental bilingual programs in 2010, and the abolition of 6th grade Spanish TAKS in 2009. The long-term implications of these relatively recent policy changes remain to be seen. However, current findings appear to support the move away from language programming as a whole towards mainstream English-only instruction.

Another implication of the current study stems from the finding that there exist considerable deviation between theory and practice. Although language acquisition theory has established the need for extended bilingual education for up to eight years in order for students to develop adequate CALP skills (Cummins, 1981; Thomas & Collier, 1997, 2002), in practice, around 60% of ELLs in the current study exited language programming between 3rd and 5th grades, with 26% of ELLs exiting in 3rd grade. The reality of educational policy is that programming decisions hinge on a multitude of complex factors beyond theory. This drives the need for more longitudinal research

examining existing student samples, in order to accurately assess actual student outcomes over time.

Current findings support that there is a lack of consensus on policy and practices across the state, which reflects the lack of educational policy standardization across the country. While there have been attempts to standardize curriculum through the establishment of The Common Core State Standards Initiative (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010), Texas is not among the 43 states to adopt these standards. Recent national educational policy changes move states further away from standardization. The U.S. Senate passed The Every Child Achieves Act (ECAA) in July 2015. This bill serves to end the federal test-based accountability system of NCLB (2002). While ECAA maintains high-stakes standardized testing as markers of progress, it will allow states more power to determine the benchmarks of adequate school performance along with appropriate next steps for addressing school underperformance. This act replaces the Adequate Yearly Progress requirement under NCLB and provides states with more power to decide how and when to change low-performing schools. Title III of ECAA specifically addresses ELLs, requiring states to measure district progress in helping ELLs meet standards, and support those for which language instruction are not effective. The bill affirms each state's responsibility to establish and implement statewide procedures for entrance and exit into ELL programs (Senate Committee on Health, Education, Labor and Pensions, 2015). As of August 2015, this bill is currently under review by the House of Representatives. While ECAA appears to be a welcomed relief from the rigid guidelines of NCLB, it

maintains the current emphasis on high-stakes testing and supports silos of educational policy in states and districts. The long-term effects of this new bill remain to be seen.

Limitations of Study

This study sought to examine the long-term academic achievement of a cohort of demographically similar students as they progressed through the educational system. The longitudinal cohort design with strict sampling criteria across all major urban school districts in Texas provided a unique perspective into the differential achievement outcomes for a specific subset of the student population. A drawback of this research design is the limited generalizability of findings. Considering the multitude of inclusion sampling criteria, results cannot be generalized to the greater student population as a whole, or even to the general ELL population in Texas.

The exclusion of students who received special education is a significant limitation of this study. Descriptive data show that 28.6% of all long-term ELL students in the sample who met demographic criteria were excluded from this study due to special education enrollment. Out of the non-ELL population, 19.7% received special education, while only 4.7% of former-ELLs received special education. These statistics demonstrate that a larger proportion of both long-term ELLs and low-income, Latino, non-ELLs receive special education at some point in their education compared to the general student population of around 10% for any given year in Texas between 1999-2010 (Texas Education Agency, 2009). By not including these students in the current study, sampling procedures may have inadvertently selected students by ability level. This limitation was also discussed above in key findings for the no services group. Along the same vein, the

exclusion of students who were retained from any grade poses another problem in the current study. By only including students who progressed to the following grade each year, this study further parsed out underachieving students. The exclusion of these two groups of exceptional students further limits the generalizability of this study.

The use of free or reduced price lunch as a proxy for poverty is another limitation of the current study. This measure is commonly used in educational research to identify low-income students and schools; however, this may not be an adequately sensitive measure of actual student poverty levels. In 2012, just over half of public school children were eligible for free or reduced price lunches, when the actual poverty rate of public school children was 22% (Snyder & Musu-Gillette, 2015). There exists much variation in the income levels of students who receive free or reduced price lunch with some poor being poorer than others. For example, a student from a household with an income at or below 130% of the poverty level is eligible for free lunch, whereas a student from a household with income between 130% to 185% of the poverty threshold is eligible for reduced lunch (Snyder & Musu-Gillette, 2015). This measure is also not static over time. In order to maintain as many students in the sample as possible, the current study included students who received free or reduced price lunch at the start of the study but did not control for possible changes in this measure across study years, which is an additional limitation to this study.

The original proposed study aimed to compare achievement outcomes of ELL students to their native English-speaking (NES) peers. Due to unexpected limitations in data availability, the first year of student demographic data available to the researcher

was 2003-2004, when study students were in 2nd grade. It was not possible to deduce student ELL status prior to this year; therefore, the current study focused on comparisons between ELL and non-ELL students as of 2nd grade, rather than NES peers. The majority of the existing literature examines achievement differences between ELLs and NES peers. It is expected that NES peers were included in this study's non-ELL group; however, current results do not offer a pure comparison of achievement differences between ELLs and native speakers.

This study examined students enrolled in bilingual education as a whole, whereas subtypes of transitional bilingual, developmental bilingual, and dual-language immersion were implemented in practice. Through examining bilingual programming as a homogenous group, nuanced differences between subprograms were lost, specifically possible differences related to length of programming received. The PEIMS database did not begin coding bilingual and ESL programming into specific subgroups until the 2009-2010 school year, which made it impossible to examine this factor at the start of the current study.

The statistical analyses utilized in this study may not have been the most sensitive for the available data. Ideally, repeated measure MANOVA procedures would have been more appropriate as these analyses take into account the correlation between the dependent variables of math and reading scores and that the same students produce scores across time points. An even better statistical approach would be the use of hierarchical linear modeling with enough school districts represented to provide adequate statistical

power at the district-level. This approach takes into consideration the nested structure of student data and statistically corrects for clustering effects.

Finally, after examining the bivariate comparisons between the sample analyzed at each time point and the original study sample, the pattern of missing data reveals that missing cases did not occur at random, but rather follow a pattern along several dimensions. Missing cases from the original sample at 7th grade were more likely to be ELL, male, and receive free lunch. The mean scores on 7th grade math and reading outcomes for missing cases between 7th and 9th grade were significantly lower than the non-missing sample. These results suggest that students from the missing sample were more likely to possess traits associated with low achievement such as higher poverty levels, and lower scores at earlier time points. It is possible that these are also students with greater school mobility who moved away from districts of interest, or dropped out of school. This poses a significant limitation, as it appears that students at higher risk for academic failure were more likely to be removed from the study. Similar to the previously described limitation of not including students who received special education, by not examining data from this missing sample, results from this study may represent a higher achieving group of students than what would naturally occur.

Researchers address missing data in multiple ways, with the "modern" approaches of multiple imputation and maximum likelihood methods being preferable to "traditional" options, including listwise deletion (Cox, McIntosh, Reason, Terenzini, 2014). Although beyond the scope of this dissertation, multiple imputation is an attractive statistical solution for addressing the problem of missing values in datasets (Rubin, 1987; Schafer,

1999). This technique involves filling in the missing entries of the incomplete dataset with values across a distribution, multiple times, such that one would have ten or more complete datasets after the imputation process. Each completed dataset is analyzed individually with the results pooled into a final result using "Rubin's Rule." The statistical inferences based on this set of final results will be statistically valid (Rubin, 1987).

A review of the literature on missing data in educational research reveals that this is a common problem in the field, especially when large-scale datasets are used (Cheema, 2014; Cox et al., 2014; Peugh & Enders, 2004). In fact, the presence of missing data on one or more variables for a proportion of the sample has become the rule rather than the exception in much of educational research (Cheema, 2014). This problem reflects the broader challenge of secondary data analyses, especially when studying a large, naturally occurring sample, where attrition is expected, such as the case with longitudinal educational research. Negative implications of missing data include a decrease in statistical power, and bias in sample statistics (Cheema, 2014). Peugh and Enders (2004) examined 389 studies with missing data published in education journals between 1999 and 2003 and found that all but six had either ignored the problem entirely, or utilized "traditional" methods of missing-data adjustments, such as listwise deletion. This highlights a prevalent methodological limitation in educational research as a whole, as well as in the current study.

Future Directions

The findings of this study begin to shed light on the long-term patterns of achievement for ELLs in Texas since the implementation of high-stakes testing accountability and accompanying policy. This study only captured a small proportion of the Texas ELL population and more longitudinal research should be conducted examining this demographically similar group of students across the state. Specifically, it would be interesting to see if current patterns of achievement hold true in suburban and rural school districts. From a policy standpoint, more information about the different types of language programming that students actually receive across the state has the potential to drive programmatic changes either towards or away from standardization.

Current findings suggest that students who received no services should be further studied. This unexpected finding is not supported by prior research and warrants additional investigation. Specifically, factors that might contribute to the relatively high performance of this group should be studied, such as individual student ability (i.e., cognitive and achievement abilities, motivations and attitudes), classroom factors (i.e., teacher training and satisfaction levels, classroom resources, curriculum opportunities, class-wide achievement levels), and composition of classroom peers (i.e., peer income levels, peer diversity). Through increased understanding of the factors that contribute to these students' achievement, educators and policymakers may be better able to support other groups of ELLs.

Future research should examine the duration of language programming and grade of program exit in relation to the amount of time needed for ELLs to be reclassified as

English proficient. These are factors often included in the existing literature, with prior research supporting longer length of programming as correlated with higher achievement outcomes (Collier, 1987; de Jong 2004; Dixon, 2012; Goldenberg 2008; Thomas & Collier, 2002). Given the findings of this study, it is unclear if this correlation would be found with the current cohort of students. The original proposed study aimed to examine length of language programming and grade of exit as potential moderators of achievement; however, they were eventually dropped from the analyses as these factors were too correlated with other study variables.

Finally, additional research is needed on the differential performance between school districts. Preliminary findings of this study suggest that some districts with higher proportion of non-ELLs have better achievement outcomes, whereas higher proportions of long-term ELLs are associated with lower achievement scores. What does the proportion of long-term ELLs in any given district say about the "health" of that district's ELL programming? What are the district factors that limit or promote the ability of ELLs to gain adequate English proficiency? These are all questions that may be further explored in hopes of improving the achievement outcomes of all learners.

Conclusions

This study provides support for the existence of differential achievement outcomes during secondary years between ELLs who received different types of language services during elementary years. There are also significant achievement differences between ELLs who are able to gain adequate English language proficiency within normal limits, and those who take longer. It appears that students who did not

receive any specialized language services have higher achievement outcomes in both math and reading than students who did receive language programming. Out of students who received language services, ESL shows some promise over bilingual education in the short-term; however, achievement differences between these groups level out by the time students enter high school.

This study is an important addition to the field of educational research as it followed a cohort of students across geographical regions in Texas over the course of seven years. The large sample size of this study promotes the generalizability of results, and more importantly, advocates for additional, future research in this area. If it is truly the case that specialized language-programming efforts are not only ineffective over the long run, but actually produce poorer results than no interventions at all, then important educational policy implications are at stake. Study findings lend support for Texas's current policy shifts away from bilingual education towards more mainstream English programming. This current study provides a strong foundation for further large-scale longitudinal investigation into ELL programming and achievement outcomes in Texas.

Appendix A

Texas ERC Proposal

February 2014

Proposed Project:
Academic Achievement Outcomes of Latino English-Language-Learners in Texas: A Longitudinal Analysis

---

Han Ren
The University of Texas at Austin
Stephanie W. Cawthon
The University of Texas at Austin

Abstract

Texas has one of the highest populations of English Language Learners (ELLs) in the U.S., with a complex system for ELL identification, program placement, and high-stakes assessment. Spanish-speaking Latino ELLs represent a large proportion of this population in this state. The long-term academic achievement of ELLs identified in elementary grades and educated in different program placements is not well known. Using growth curve analysis within a hierarchical linear modeling framework, this study compares student achievement, as measured by high-stakes testing, of Latino former-ELLs and long-term ELLs enrolled in Bilingual Education, English as a Second Language, and their Native English Speaking demographically similar peers. Achievement data from one district will be examined at four time points spanning 3rd-9th grade. Findings will suggest implications of Texas educational policy on ELL student achievement.

Reader's Guide

Research Questions

(1) Compared with native-English speaking peers with similar demographics (i.e., Latino, low-SES, and district of attendance), what are the effects of program enrollment in elementary grades and time of program exit (ESL vs. early-exit bilingual vs. late-exit bilingual) on ELL student academic achievement in secondary grades?

(2) How do former-ELLs (defined as reclassified as English proficient in 5th grade or earlier) perform on measures of academic achievement from 3rd to 9th grade compared to long-term ELLs (defined as reclassified to English proficient after 5th grade, or never reclassified) and their Native-English-Speaker counterparts?

(3) Assuming that hypothesis 2 is confirmed, and former-ELLs have higher achievement outcomes than long-term ELLs in 9th grade, what individual differences between groups significantly contribute to differential achievement outcomes?

Data Elements from ERC Database

This project will require the following variables from ERC data for the years 2003-2004; 2004-2005; 2005-2006; 2006-2007; 2007-2008; 2008-2009; 2009-2010:
TEA – Enrollment
TEA – Attendance
TEA – Assessment
(Specific variables for each data set and descriptions of those variables may be found of pages 6-7 of the full proposal.)

Research Methods

Using TEA data spanning from the 2003-2010, this study will retrospectively investigate the effects of available classroom program models for two groups of students who are identified as ELL in elementary school on their long-term academic achievement. Achievement data as measured by performance on the TAKS from a cohort of approximately 2000 participants (500 ESL students, 500 early-exit bilingual education students, 500 late-exit bilingual education students and 500 comparison native English speaking students) will be examined. Independent variables (i.e. immigrant status, age of U.S. school enrollment, age of reclassification as English proficient, beginning English proficiency level, primary language model received, length of program enrollment, grade of program exit) that have been shown to be significant within the literature as factors affecting ELL achievement will similarly be examined. To control for the effects of diverse backgrounds, only Spanish-speaking immigrant ELLs from low-income families will be included in the target groups. The comparison group will be comprised of low-income, native English speaking, Latino students. Analysis for the data will include descriptive information at all levels, growth curve analyses within a hierarchical linear modeling framework, and multiple regression modeling of independent variables. *(A full description of the project methodology may be found in the proposal on pages 4-7; these*

pages cover in-depth the participant characteristics, data set, sampling procedures, and a table of variables/years of data needed.)

Researcher's Qualifications

Han Ren is a fifth year doctoral candidate in the School Psychology program at the University of Texas at Austin. She is conducting this study as her dissertation, in partial fulfillment of the requirements for the degree of doctor of philosophy. Han has prior research experience working with large databases and has taken classes on experimental design, regression methods, and multi-level modeling. Dr. Stephanie Cawthon is an associate professor in the department of educational psychology. Her research interests focus on investigating issues of equity and access in education from multiple vantage points. She serves as Han's dissertation committee chairperson for the purposes of this study.

This study has been processed by the Office of Research Support and determined as "Exempt" from IRB review (IRB # 2013-07-0061).

Benefit to Texas

The utilization of these data sets, combined together through common identifiers and a longitudinal structure, provides an ideal setting to utilize the full capacity of a wide spectrum State Longitudinal Data System such as the ERC as well as the study of long-term educational outcomes. Contributions to the academic realm include the expansion of research connecting elementary language programming models to high school achievement outcomes as measured by high-stakes standardized tests. This research will provide useful and meaningful information on the long-term impacts of Texas policies since No Child Left Behind, related to the identification and education of ELLs and their future achievement outcomes. These findings may potentially inform policy in Texas, particularly in multicultural, urban, school districts similar to district examined in this study, and throughout the nation on the identification and education of ELLs. *(Detailed reasoning behind the project and the potential policy implications of this project are available in the full proposal on pages 1-3 and 8.)*

Related Policy

There are many potential policy implications resulting from this study regarding the identification, education, and assessment of ELLs. This study may provide empirical support for specific models of language programming, language of instruction, and duration of language supports for ELLs. Results may inform assessment decisions for provision of language-specific high-stakes standardized tests for ELLs beyond what is currently available in Texas. Finally, the identification of individual characteristics that may delineate high-achieving former-ELLs from their lower-achieving long-term ELLs may provide policy implications for earlier educational interventions for students at-risk for becoming long-term ELLs.

Dissemination Strategy

Han Ren and Dr. Stephanie Cawthon will work to complete a policy brief with its findings that discusses the implications of long-term ELL achievement outcomes based on elementary language program models of instruction. The researchers will conduct broader academic connections, including presentations at conferences and potential for peer review journal submission when sufficient results have been obtained.

Financial Resources

This project is not funded. Personal funds will be utilized for this study.

FULL PROPOSAL ATTACHED

Introduction

English Language Learners (ELLs) are among the fastest growing student populations in the U.S. public school system, with Texas ranking second highest of all states for the number of students attaining English proficiency (U.S. Department of Education, 2006). The growth of ELLs has been increasing at a faster rate compared to overall student enrollment in recent years. Between the 1995-96 and 2005-06 school years, total enrollment in the state of Texas increased 13.6% while ELL enrollment increased 34.0% (National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, 2007). Among this diverse group of ELLs, Latino students represent the fastest rising group, comprising one out of every five U.S. public school students in 2008 (Fry & Gonzales, 2008). As a state with one of the highest numbers of ELLs, Texas bears strong responsibility to establish effective policies in the identification, education, and assessment of these students. As ELL populations continue to grow within public school systems, other states will look to Texas as an exemplar for ELL education models. The need to identify, educate and assess the growing population of ELL students across the nation presents a formidable challenge for researchers, policy makers and educators alike.

The issue of ELL education has been a prominent research and policy topic both historically and more recently within the context of the national education reform movement. The history of bilingual education is long and contentious, marked by litigation by minority groups in the pursuit of educational equality (San Miguel, 2004; Valencia, 2008). Contemporary bilingual education policy was established in the 1960's, with the passage of the Bilingual Education Act of 1968. This act coupled with the Equal Educational Opportunity Act (EEOA) of 1974 paved the way for federal legislation enforcing equitable and differentiated education for language minority students (Stewner-Manzanares, 1988; Valencia, 2008). Supporters of bilingual education consider language equality to be a civil rights issue, with appropriate educational access for ELLs as a vital component for achieving a multicultural America (San Miguel, 2004). Conversely, bilingual education has been met with much opposition from English-only education proponents, who fear that language education will divide America into a bicultural and bilingual society, undermining the country's foundational Anglo-Protestant roots (Hempel, Dowling, Boardman, & Ellison, 2012; Huntington, 2004). This sociopolitical debate continues to play out in recent educational reform policies.

On a policy level, the No Child Left Behind Act of 2001 (NCLB) essentially eliminated the federal provision of bilingual education by replacing it with English-only legislation in order to meet the linguistic needs of ELLs (San Miguel, 2004). While local and state educational agencies may still offer bilingual programming if perceived to be appropriate, federal funding incentives place emphasis on ELLs attaining English proficiency as quickly as possible, promoting the English-only educational agenda. Additionally, the United States Department of Education allows each state to devise their own rules and regulations surrounding the identification, education, and assessment of ELLs, resulting in inconsistent laws and practices across states. NCLB (2002) requires all

students to participate in high-stakes testing as part of a federal accountability system, whereas ELL data were formerly excluded from school-wide score reports. This requirement places the onus on schools to properly educate ELLs and ensure they are receiving an equitable education as their Native English Speaker (NES) peers, as federal funding are now contingent on all student test scores in English.

While federal policy aims to ensure all students receive an equitable education, the amalgamation of these policies in practice has yielded dismal achievement outcomes for ELLs. The achievement gap between ELLs and their NES peers is well established (U.S. Department of Education, 2009; 2010; 2012a; 2012b; 2012c), as illustrated through higher dropout rates, lower academic achievement, and an overrepresentation of ELLs in special education programs compared to their NES peers. There is a significantly lower proportion of ELLs who pass statewide high-stakes tests under NCLB (2002) compared to their NES peers, especially at the secondary level (Smith, 2010; TEA, 2008). This achievement gap has sparked researchers to examine the root causes of this educational inequity and track the progress of current and former ELLs as they navigate school systems, in order to develop and propose viable policy solutions.

At the center of this research are the development and evaluation of effective instructional programming for ELLs. Taking into consideration culture, context, policy and second language acquisition theory, scholars aim to identify ELL classroom models with the most promising long-term achievement outcomes. Instructional programming for ELLs is in a state of constant evolution, with the identification of six or more program models, ranging from dual-language bilingual education to English-only structured immersion classrooms (Thomas & Collier, 2002; Genesee, 1999). The availability of different program models varies significantly among states and school districts, resulting in inconsistent provision of services. However, the most prevalent language instruction program models can be reduced to three categories: mainstream English-only, Bilingual Education, and English as a Second Language (ESL) (Genesee, 1999). The establishment of the most effective instructional model is a hotly contested research topic, especially given the constraints of shifting educational policies (Thomas & Collier, 2002). While some scholars identify bilingual education as producing the most promising achievement outcomes for ELLs (August & Shanahan, 2006; Genesee, Lindholm-Leary, Saunders, & Christian, 2006; Goldenberg, 2008; Mora, Wink, & Wink, 2001; Thomas & Collier, 1997; 2002), other studies fail to find clear advantages between language instruction models (Galvan-Luis, 2010; Nakamoto, Lindsey, & Manis, 2012; Slavin, Madden, Calderon, Chamberlain, & Hennessy, 2011). There is a lack of consensus on what constitutes best practice for the identification, education, and assessment of ELLs (Abedi, 2004; August & Hakuta, 1997). However, based on the achievement gap, it is clear that current educational systems are not adequately meeting their academic needs.

Research on current ELLs has focused on achievement in high school, examining the impacts of accountability, policy, and language on dropout rates and academic achievement on Spanish-speaking students primarily from low-socioeconomic status

(SES) backgrounds within urban school districts (McNeil, Coppola & Vasquez-Heilig, 2008; Vasquez-Heilig & Darling-Hammond, 2008). Through this research, a group of unique ELLs has emerged termed "long-term ELLs," who require five or more years to attain English proficiency (Menken & Kleyn, 2009; Olsen, 2010). This subset of ELLs faces exceptional challenges, with poor achievement outcomes. They are frequently adolescents who appear orally bilingual, yet lack academic literacy skills in either their native language or English (Menken & Kleyn, 2009; Olsen, 2010). These students are included in the current study in hopes of understanding what factors allow some ELLs to achieve English proficiency, while others remain perennially stuck in ELL classification.

Although there is considerable evidence that current ELLs have significantly lower academic achievement than NES students, the literature on long-term achievement outcomes for students who have been reclassified from ELL to English language proficient is just beginning to emerge. A growing body of recent research supports positive long-term achievement outcomes for these former-ELLs (Ardasheva, Tretter, & Kinny, 2012; Flores, Batalova, & Fix, 2012; Halle, Hair, Wandnar, McNamara, & Chien 2012; Kim, Herman, & National Center for Research on Evaluation, 2009; Kim, Herman, & National Center for Research on Evaluation, 2012). Factors that contribute to favorable former-ELL achievement outcomes include maintaining grade-level academic proficiency, three years of language programming, and earlier age of reclassification (Flores, Batalova, & Fix, 2012; Halle et al., 2012).

The literature comparing former-ELLs who receive differential instruction is still lacking. The research that does compare student achievement outcomes post-reclassification from different program models (de Jong, 2004; Kim & Herman, 2009; Thomas & Collier, 2002) rarely follows students past middle school. Most of these studies also utilized data collected before the implementation of NCLB (2002), which has significantly altered the academic environment, as ELLs' test scores now contribute to federal funding decisions. Furthermore, the ELL populations featured in prior studies come from various backgrounds, differing in native language, SES, state of enrollment, and geographic setting (rural versus urban). Thus it is difficult to draw conclusions on the long-term academic trajectories of these former-ELLs with diverse backgrounds, and there may be a cohort effect as a result of changing educational policies. Further understanding of a cohort of ELLs as they mature and progress through the educational system will provide a clearer understanding about the academic trajectories of ELLs, and the relative effectiveness of various elementary language instruction models on future achievement.

This study will retrospectively examine student data from a large urban Texas school district to investigate the effects of available classroom program models on future academic achievement for two groups of demographically similar ELLs identified in elementary school. The achievement trajectories of long-term ELLs and former-ELLs compared to their NES peers will be compared. This cohort of students is an important population of study due to the lack of demographically controlled, longitudinal analysis

on the long-term academic achievement of ELLs in previous research. Through examining this data, one will be able to infer the effectiveness of recent available academic models for preparing ELLs to exit language programming and enter mainstream education. The contributing factors and achievement differences between long-term ELLs and former-ELLs will be reviewed in order to assess why some ELLs are able to attain English proficiency while others struggle.

In particular, this study will evaluate student high-stakes test score differences between ESL and bilingual education program models, as well as in comparison to English-only mainstream education. Due to the variety and inconsistency of language programming availability, the conglomerate categories of ESL and bilingual education will be used to maximize statistical power, and to simplify analyses. To control for the effects of diverse backgrounds, only Spanish-speaking ELLs from low-SES families will be included in the target groups. The comparison group will be comprised of low-SES, native English speaking, Latino students. Student-level demographic data will be used to determine students SES, native language, and ethnicity. Using growth curve analysis within a hierarchical linear modeling framework, this study will compare the high-stakes test scores of two groups of ELLs enrolled in two language program models and mainstream English-only instruction at four points in their education, from 3rd through 9th grade, in order to compare their relative implications for student achievement.

Research Questions

(1) Compared with native-English speaking peers with similar demographics (i.e., Latino, low-SES, and district of attendance), what are the effects of program enrollment in elementary grades and time of program exit (ESL vs. early-exit bilingual vs. late-exit bilingual) on ELL student academic achievement in secondary grades?
(2) How do former-ELLs (defined as reclassified as English proficient in 5th grade or earlier) perform on measures of academic achievement from 3rd to 9th grade compared to long-term ELLs (defined as reclassified to English proficient after 5th grade, or never reclassified) and their Native-English-Speaker counterparts?
(3) Assuming that hypothesis 2 is confirmed, and former-ELLs have higher achievement outcomes than long-term ELLs in 9th grade, what individual differences between groups significantly contribute to differential achievement outcomes?

Methods

Participants

Data set. This study will involve a secondary data analysis of individual student-level data derived from Texas Education Agency's Public Education Information Management System (PEIMS). Student data from 2003-2010 will be collected from an urban, multicultural Texas school district where the language programming models of interest are widely available.

The PEIMS is a statewide database supervised by the Texas Education Agency

(TEA), and encompasses all data requested and received by TEA about public education in Texas. The database includes individual student information on: enrollment, demographics, special program participation, attendance, discipline, and school leaver data such as withdrawn students, graduates, and dropouts. Additionally, the PEIMS is linked to a separate database containing student results on the Texas Assessment of Knowledge and Skills (TAKS), the statewide high-stakes assessment.

Sampling. The current research will be a retrospective longitudinal study examining student data from approximately 2000 students within one school district spanning 3rd grade through 9th grade between the 2003-2004 and 2009-2010 school years. To control for the clustering effects of school enrollment, only students from three high school feeder patterns will be included. To address the first research question, the ELLs will be randomly selected from three program models, with 500 students from each: ESL, early-exit bilingual education, and late-exit bilingual education In addition, data from a comparison group of 500 NES students matched on demographic variables of Latino and low-SES enrolled in mainstream English classrooms within the same three high school feeder patterns in one school district will also be examined. From this existing sample, ELL participants will be re-categorized into two groups to address the remaining research questions: former-ELLs who started school as ELL but were reclassified as English proficient before the start of secondary education (6th grade), and long-term ELLs who also started school as ELL but were either reclassified as English proficient after 5th grade, or remained ELL for the duration of their schooling, with an English proficiency level below Advanced on the TELPAS. All student data will be obtained from the same database.

Student sampling will be carefully conducted to only include data from those who meet the inclusion criteria, due to the diversity of ELLs in Texas. For the three target groups of ESL and early/late-exit bilingual education, students must attend school in the district by 3rd grade, be identified as ELL, and enrolled in one of the three language-program models. ELL status is determined on an individual basis by the Language Proficiency Assessment Committee (LPAC) on each campus (Texas Education Agency, 2012). Students who qualify as ELL have a home language that is not English, and a beginning TELPAS score. The LPAC also determines when students are exited from ELL status, and become former-ELL students. This is achieved after three years of schooling in the U.S. for immigrant ELLs or a TELPAS score of advanced. For the comparison group, students must be enrolled in mainstream English classrooms, and not be categorized as ELL at any point in their education.

All students must remain in their instructional program model (ESL, early/late-exit bilingual) for at least two years, through the end of 4th grade, and only be enrolled in one program model. Additionally, target students must meet the following demographic criteria: Latino, receiving free or reduced meals, and come from Spanish speaking households as indicated by the Home Language Survey (HLS). Students in the comparison group must meet all the demographic criteria except for the HLS, which must

be English. The comparison group will be matched on the variables of race, grade, and free or reduced lunch. All students must also remain within the school district for the duration of the study, through their 9th grade years, although mobility between schools within the same high school feeder pattern is allowable, and students who drop out of school in 9th grade will still be included. In order to control for attrition due to student mobility and dropout, only students who are enrolled in the district within the same high school feeder pattern for the duration of the study will be included. Additional exclusion criteria are enrollment in special education and students with two or more missing data points, as this study aims to only examine ELLs educated in general education settings and produce as complete of a data set as possible. Students who do not meet these requirements will not be included in the data sample. Given the large number of available ELLs from which this sample draws upon, these rigorous sampling procedures are not anticipated to be problematic.

Variables

The dependent variables for all research questions will be measures of academic achievement, as represented by TAKS scores in reading and math. For the first two research questions, TAKS scores will be examined in 3rd, 4th, 7th, and 9th grade. For the third research question, the dependent variable will be achievement measured in 9th grade only. Due to limited data availability, English language proficiency (ELP) at baseline only will be included as an independent variable, and measured through student TELPAS scores, a statewide assessment that is administered annually to ELLs. Research question 3 examines seven additional individual-level independent variables described within the literature as relevant predictors of ELL achievement. These independent variables aim to answer the following questions:

(1) Is the target student an immigrant? As indicated by the "immigrant" variable of not being born in the U.S. and have not been attending one or more schools in any state for more than three full years.
(2) At what grade level did target students exit their language programming and enter mainstream English classrooms?
(3) How many years were target students enrolled in their language program model?
(4) What was the primary classroom model of enrollment for target students?
(5) At what age (if any) were target students reclassified as English language proficient?
(6) At what age did target students begin attending school in the U.S.?
(7) What was the target student's baseline English proficiency level?

Variables To Be Collected at Each Time Point

3rd Grade: 2004	**4th Grade: 2005**	**7th Grade: 2008**	**9th Grade: 2010**	**Additional Independent Variables:**
TAKS reading score	TAKS reading score	TAKS reading score	TAKS reading score	Immigrant Status – Yes/No

TAKS math score	TAKS math score	TAKS math score	TAKS math score	Grade of exit from ESL/BE
TELPAS score				Length of enrollment in ESL /BE
				Primary language model received
				Age of ELP reclassification
				Age of U.S. school enrollment

Data Elements from ERC Database

Student demographic information (student identifier, race/ethnicity, free/reduced lunch, LEP indicator code, home language, immigrant status):
p_enroll_demog(03); p_enroll_demog(04); p_enroll_demog(05); p_enroll_demog(06); p_enroll_demog(07); p_enroll_demog(08); p_enroll_demog(09); p_enroll_demog(10)

Information on the setting used in providing instruction to students (ESL program type code, Bilingual program type code, no language programming) derived from attendance data:
p_attend_demog(03); p_attend_demog (04); p_attend_demog (05); p_attend_demog (06); p_attend_demog (07); p_attend_demog (08); p_attend_demog (09); p_attend_demog (10)

Parent/guardian approved placement in required bilingual or ESL programs:
p_lep_permission(03)

TAKS information:
taks3sp_march_fy04; taks3sp_april_fy04; taks3sp_july_fy04; taks3_march_fy04; taks3_april_fy04; taks3_july_fy04; taks4_fy05; taks4sp_fy05; taks7_apr08; taks9_apr10
TELPAS information:
telpas_sp04

Statistical Analyses

This study will employ a growth curve analysis using a hierarchical linear modeling (HLM) framework, to study the change of academic achievement and related outcomes between target and comparison students. The goal of growth curve modeling is to determine how the growth of an outcome variable depends on individual

characteristics. In this model, the data structure will consist of individual observations nested within the subject across time, where variation in final status and growth rates are considered random factors, serving as targets of inference.

Determination of sample size and statistical power analysis. This study will employ the general framework of power analysis (Cohen, 1988) and the framework for optimal design of longitudinal study using HLM (Raudenbush & Liu, 2001) to determine the needed sample size. Specifically, a power analysis for models using HLM aims to determine the number of required subjects for this study that produces statistical significance ($\alpha = 0.05$) with adequate statistical power ($[1 - \beta] \geq 0.80$, where β is the probability of making a type II error). For an HLM analysis, factors affecting needed sample size include: the effect size (i.e. the difference on an outcome variable between the target and comparison groups), intra-class correlation coefficient (i.e. percentage of variation on outcome variable that is due to students), and number of time points the longitudinal data provide (four for this study). The power analysis will be conducted after preliminary data are available. It is expected that the estimated sample size of 2000 students with 500 subjects per category will result in sufficient power to reject a false hypothesis. The study will employ the Optimal Design software (Spybrook, Randenbush, Cogdon, & Martinez, 2009) to perform this power analysis.

Tests of research questions. Prior to the primary analyses, it is necessary to confirm the data structure in order to determine whether this study will require a two-level or three-level HLM model. The primary advantage of the HLM approach is to correct for clustering effects to ensure a rigorous and unbiased statistical inference. There may be two kinds of nesting or clustering effects existing in the current dataset: students' observations at different time points may be nested within students and students may be nested within schools. If only the first clustering exists, a two-level HLM is adequate; if both types of clustering effects exist, then a three-level HLM is necessary. In general, I will employ a two-level HLM growth curve model to test the first two research hypotheses. All analyses specify a linear individual growth model, as the limited number of observations per individual (i.e., four time points) does not allow for a curvilinear model (Raudenbush & Bryk, 2002). The last research question will employ a linear multiple regression model using wave-four outcome data and the collected independent variables for each target student. Preliminary analysis will be conducted to obtain descriptive statistics about the data, including the variance of observations within individuals, the variance of observations between individuals, the total variance, and the proportion of this variance that is between subjects.

Implications

There are important implications for educational policy in Texas if significant results are found in this study. If former-ELL and long-term ELL students are indeed performing lower academically than their NES peers, while controlling for race and income, then this suggests that the current educational systems are not adequately

meeting ELLs' needs. A longitudinal perspective on student achievement after the consequential adoption of NCLB (2002) would provide a significant evaluation on how this policy change is impacting ELLs in practice. By shifting the focus from current ELL to former-ELL students, we can examine the long-term impacts of language programming models. Furthermore, if students enrolled in bilingual education perform higher than those enrolled in ESL programming, then this would support the value of developing native language proficiency to aide in the acquisition of English, thus confirming the theories of second language development. This finding would also reinforce Ruiz's (1984) "language as a right" and "language as a resource" perspectives as having better student outcomes for both language and culture development than the more prevalent "language as a problem" framework.

On a policy level, given the theory that it takes four to eight years to develop academic English proficiency, the length of the language "grace period" for new immigrants in Texas should be extended from three years to at least four. If this study supports a longer duration of language instructional programming, and later transition to mainstream English classrooms as having better student outcomes, then this could have a significant impact on policy decisions. This finding would support prolonged enrollment in bilingual education, and encourage the availability of both bilingual and ESL programming in secondary education.

Overall, the findings from this study could encourage policy makers to critically examine their decisions, and evaluate whether current educational policies are grounded in theory and adequately supporting ELLs' success, or merely arbitrary decisions of popularity and convenience. This study could potentially shed light on the real-world impacts of current language program models, and provide empirical evidence to fuel future policy decisions to better support the nation's growing ELL populations.

Appendix B

Proposed Hierarchical Linear Modeling Methodology

This study will employ a growth curve analysis using a hierarchical linear modeling (HLM) framework, to study the change of academic achievement and related outcomes between target and comparison students. The goal of growth curve modeling is to determine how the growth of an outcome variable depends on individual characteristics. In this model, the data structure will consist of individual observations nested within the subject across time, where variation in final status and growth rates are considered random factors, serving as targets of inference.

Compared to conventional statistical methods such as a within-subject design analysis of variance (ANOVA) and a repeated-measure multivariate analysis of variance (MANOVA), HLM has several advantages for the analysis of longitudinal data. First, responses on any outcome variable from the same student over time will be correlated, thus violating the assumption of independent observations, which is embedded in most statistical models dealing with cross-sectional data. HLM takes this intra-class correlation (ICC) into consideration. The ICC measures the proportion of variance in an outcome variable that lies between students, thus a high ICC indicates that much variation in the outcome variable is due to factors within students, which signifies the need to use HLM. Second, when applying conventional linear models (e.g. an ordinary least square regression or OLS) to analyze longitudinal data, one generally underestimates the standard errors of the regression coefficients and therefore may erroneously infer statistical significance. HLM effectively addresses this problem as well as others inherent

in longitudinal data, such as unequal groups at each data point, missing data across time points, varying time intervals between repeated measures, and the need to control for the effects of potentially confounding independent variables (Diggle, Liang, & Zeger, 1994; Lindsey, 1993; Raudenbush & Bryk, 2002). These advantages make HLM more appropriate than conventional repeated measures analyses used in longitudinal studies.

Determination of sample size and statistical power analysis. This study will employ the general framework of power analysis (Cohen, 1988) and the framework for optimal design of longitudinal study using HLM (Raudenbush & Liu, 2001) to determine the needed sample size. Specifically, a power analysis for models using HLM aims to determine the number of required subjects for this study that produces statistical significance ($\alpha = 0.05$) with adequate statistical power ($[1 - \beta] \geq 0.80$, where β is the probability of making a type II error). For an HLM analysis, factors affecting needed sample size include: the effect size (i.e. the difference on an outcome variable between the target and comparison groups), intra-class correlation coefficient (i.e. percentage of variation on outcome variable that is due to students), and number of time points the longitudinal data provide (four for this study). The power analysis will be conducted after preliminary data are available. It is expected that the estimated sample size of 2000 students with 500 subjects per category will result in sufficient power to reject a false hypothesis. The study will employ the Optimal Design software (Spybrook, Randenbush, Cogdon, & Martinez, 2009) to perform this power analysis.

Tests of research questions. Prior to the primary analyses, it is necessary to confirm the data structure in order to determine whether this study will require a two-

level or three-level HLM model. As noted earlier, the primary advantage of the HLM approach is to correct for clustering effects to ensure a rigorous and unbiased statistical inference. There may be two kinds of nesting or clustering effects existing in the current dataset: students' observations at different time points may be nested within students and students may be nested within schools. If only the first clustering exists, a two-level HLM is adequate; if both types of clustering effects exist, then a three-level HLM is necessary. For the first two research questions, one-way ANOVA with random effects will be conducted to examine the ICC between time and students and the ICC between individual students and schools using the baseline data. Tests of the clustering effects will be based on the ICC. If the ICC between time and students is higher than 0.2, this study will necessitate the use of a two-level HLM. If both ICCs are higher than 0.2, this suggests that observations at different time points are nested within students and students are nested within schools, and the study must employ a three-level HLM. For the third research question, I will run the same one-way ANOVA with random effects analysis on the data obtained in 9^{th} grade to determine the ICC. For this analysis, if the ICC is above 0.2, research question 3 will require a two-level HLM analysis where students are nested within schools. If the ICC is below 0.2, then the analysis will be simplified to an OLS regression.

For the purposes of this proposal, it is assumed that the ICC between students and schools will be below 0.2 for the data at both baseline and final time points. In general, I will employ a two-level HLM growth curve model to test the first two research hypotheses. All analyses specify a linear individual growth model, as the limited number

of observations per individual (i.e., four time points) does not allow for a curvilinear model (Raudenbush & Bryk, 2002). Assuming the ICC between students and schools is below 0.2, the last research question will employ a linear multiple regression model using wave-four outcome data and the collected independent variables for each target student. Preliminary analysis will be conducted to obtain descriptive statistics about the data, including the variance of observations within individuals, the variance of observations between individuals, the total variance, and the proportion of this variance that is between subjects.

Research questions 1 & 2.

(1) Compared with native-English speaking peers with similar demographics of Latino, low-SES, and district of attendance; what are the effects of program enrollment in elementary grades and time of exit (ESL vs. early-exit bilingual vs. late-exit bilingual) on ELL student academic achievement in secondary grades?

(2) How do former-ELLs perform on measures of academic achievement from third to ninth grade compared to long-term ELLs and their NES counterparts?

Unconditional model. The unconditional growth model expresses the status of the outcome (TAKS scores) for an individual over time in terms of the rate of growth at level-1. At level-2, the unconditional model is considered to determine if there is true variability of the growth curves across individuals. Hypothesis testing will be conducted using the unconditional model to ensure that variability exists between individuals over time. If the null hypothesis is rejected, signifying the existence of true variation, then the conditional model will be employed.

Linear growth model equation, unconditional model:

Level-1

$Y_{ti} = \pi_{0i} + \pi_{1i}\,\alpha_{ti} + e_{ti}$

Level-2

$\pi_{0i} = \beta_{00} + r_{0i}$

$\pi_{1i} = \beta_{10} + r_{1i}$

Where:

Y_{ti} is the TAKS score at time *t* for student *i;*

π_{0i} is the expected TAKS score for student *i* at the intercept (when time = 0), which is the final data point in this study, measured in 9th grade;

π_{1i} is the linear growth rate of TAKS score for student *i* at that period of the study;

α_{ti} is the measure of time in years (i.e., 3rd, 4th, 7th, or 9th grade) for student *i*, where $\alpha_{1i} = -6$, $\alpha_{2i} = -5$, $\alpha_{3i} = -2$, and $\alpha_{4i} = 0$;

e_{ti} is the level-1 random effect, assumed to be constant across all individuals, normally distributed, with a mean of 0, and a constant variance;

β_{00} is the expected average TAKS score at the intercept or final status (time = 0), in 9th grade;

β_{10} is the slope, the expected average growth rate for TAKS score;

r_{0i} is the random effect associated with the average TAKS score for student *i*;

r_{1i} is the random effect associated with the growth rate for student *i;*

Conditional model. The level-1 equation for the conditional model remains the same as above. The level-2 model will include the addition of the explanatory variable of group enrollment. For research question 1, this group enrollment will be dummy-coded to represent the four program enrollment categories of mainstream English, ESL, early-exit bilingual education, and late-exit bilingual education. For research question 2, this group enrollment will be dummy-coded to represent the three ELL categorizations of LTELL, former-ELL, and NES.

(1) Linear growth model equation, conditional model, research question 1:

Level-2

$$\pi_{0i} = \beta_{00} + \beta_{01}W1_i + \beta_{02}W2_i + \beta_{03}W3_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}W1_i + \beta_{12}W2_i + \beta_{13}W3_i + r_{1i}$$

Where:

The reference group is mainstream English enrollment;

$W1$ is the dummy-coded student-level group predictor of ESL enrollment (W1 = 1 if ESL, W1 = 0 for all others)

W2 is the dummy-coded group predictor of early-exit bilingual education enrollment;

W3 is the dummy-coded group predictor of late-exit bilingual education enrollment;

β_{00} is the average TAKS score for the reference group at time = 0 (9^{th} grade);

$\beta_{00} + \beta_{01}$ is the average TAKS score for group W1 at 9th grade;

$\beta_{00} + \beta_{02}$ is the average TAKS score for group W2 at 9th grade;

$\beta_{00} + \beta_{03}$ is the average TAKS score for group W3 at 9th grade;

β_{10} is the expected change rate for the reference group over the four time points;

β_{11} is differential change rate of group W1 in relation to the reference group, also known as the cross-level interaction;

β_{12} is differential change rate of group W2 in relation to the reference group;

β_{13} is differential change rate of group W3 in relation to the reference group;

r_{0i} is the random effect associated with the average TAKS score of the reference group;

r_{1i} is the random effect associated with the change rate for the reference group;

(2) Linear growth model equation, conditional model, research question 2:

Level-2

$$\pi_{0i} = \beta_{00} + \beta_{01}W1_i + \beta_{02}W2_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}W1_i + \beta_{12}W2_i + r_{1i}$$

Where:

The reference group is NES;

W1 is the dummy-coded student-level group predictor of LTELL;

W2 is the dummy-coded group predictor of former-ELL;

The remaining variables are interpreted in the same way as the above conditional model for research question 1.

Research question 3. Assuming that hypothesis 2 is correct, and former-ELLs have higher achievement outcomes than long-term ELLs in 9th grade, what individual differences between groups significantly contribute to differential achievement outcomes?

Multiple regression. Assuming the ICC between students and schools is below 0.2, the analysis for this research question will employ multiple regression using data from the final time point. Growth curve modeling is not an appropriate analysis as the variables of interest are not available prior to the start of this study, thus violating the temporal order requirement of HLM that independent variables are collected before dependent variables. If the ICC between students and schools is above 0.2, the planned regression model will be replaced by a two-level HLM model similar to the regression model with the addition of random effects to control for the nesting structure of students in schools.

Prior to this analysis, a test of multicollinearity will be conducted by examining the variance inflation factor (VIF) in order to ensure that the assumption of no multicollinearity has been met. A VIF greater than 10 indicates problematic multicollinearity. If any of the VIFs reach the 10 threshold, one of these variables will be dropped from the study. Descriptive statistics will be reported for each of the significant independent variables.

To test group differences, I will employ a multiple regression model that tests a series of interaction terms. An interaction effect generally indicates that the impact of an independent variable varies by the level of another independent variable. Under the current context, if a group difference exists, a significant interaction between an independent variable and the indicator variable of group is expected. This research question will employ the following multiple regression equation:

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \beta_3 X3_i + \beta_4 X4_i + \beta_5 X5_i + \beta_6 X6_i + \beta_7 X7_i + \beta_8 G1_i + \beta_9 G2_i + \varepsilon_i$$

Where:

X1 = Immigrant status

X2 = Grade of exit from language programming

X3 = Length of enrollment in language programming

X4 = Primary language model received

X5 = Age of ELP reclassification

X6 = Age of U.S. school enrollment

X7 = TELPAS score at baseline

G1 = Dummy-coded group participation for former-ELL (G1 = 1 if former-ELL, G1 = 0 for all others)

G2 = Dummy-coded group participation for LTELL;

the reference group is NES;

ε = Residual, the effects of all other variables not included in the equation

The associated β for each variable is the effect of that variable on the predicted outcome, while holding all other variables constant.

Significance testing will be done on each of these variables. Next, the inclusion of an interaction term between group participation and each of the variables (i.e., G1*X1, G1*X2, … G1*X7, G2*X1, G2*X2, … G2*X7) will be added to the model one at a time, and individually tested for significance. This step will determine whether group differences exist for each of the variables of interest, while holding all other variables constant. After running all 14 individual models, all significant interactions ($p < 0.05$) from the individual equations will be combined into one model. Non-significant interactions from this model will be dropped. The final model will include all main effects (i.e., the seven independent variables regardless of whether or not they are significant), as well as significant interaction terms. These statistically significant interactions indicate group differences. Results from the interaction model will be presented graphically using model-based prediction. Depending on which interactions remain significant, the final model may look like the following equation:

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \beta_3 X3_i + \beta_4 X4_i + \beta_5 X5_i + \beta_6 X6_i + \beta_7 X7_i + \beta_8 G1_i + \beta_9 G2_i + \beta_{10} X1_i * G1_i + \beta_{11} X7_i * G1_i + \beta_{12} X3_i * G2_i + \beta_{13} X5_i * G2_i + \varepsilon_i$$

Glossary

Bilingual Education: Instructional programming for ELLs delivered in mix of English and native-language, taught by bilingual-certified teachers. Goal is for students to achieve academic English proficiency in 2-7 years. Degree of emphasis on bilingualism and biculturalism varies; however, most bilingual program utilize native language as a stepping stone for English acquisition.

English as a Second Language (ESL): Instructional programming for ELLs delivered in English only, taught by ESL-certified teachers. Goal is for students to achieve academic English proficiency in 1-3 years. Native language is ignored.

English-language-learner (ELL): A student in the process of acquiring English as a second language. The newer and preferred term (over LEP) as it does not ascribe deficiency to the learner.

Former-ELL: A student who began school as ELL but was reclassified as English language proficient and mainstreamed to English-only classes. In this study, these students were reclassified before the start of middle school.

Limited-English-Proficient (LEP): A term used historically and in federal documents to refer to students acquiring English. In Texas prior to 2010, LEP was used to identify immigrant ELLs who have been in the country for <3 years and eligible to receive language accommodations on statewide high-stakes testing.

Long-Term ELL: A student who began school as ELL and remained ELL for 5+ years of U.S. schooling, into secondary education. May or may not achieve English proficiency by the end of high school.

Mainstream English (No Services): Instructional programming for monolingual English-speaking students, taught by teachers who are not required to hold certification beyond content area. This is the program that ELLs are exited to after achieving English proficiency.

Native English Speaking (NES): A student whose native language is English. Demographically similar NES students are included in the non-ELL comparison group in this study.

Non-ELL: Students who were not identified as ELL. These students are used as the comparison group in this study.

Texas Assessment of Knowledge and Skills (TAKS): The high-stakes assessment of content knowledge for Texas from 2003-2011. Scores in math and reading will be used as outcome measures of achievement in this study.

Texas English Language Proficiency Assessment System (TELPAS): The statewide English Language Proficiency assessment of Texas. In this study, baseline TELPAS scores are included as a predictor of future achievement outcomes.

References

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Abedi, J. (Ed.) (2007). *English language proficiency assessment in the nation: Current status and future practice.* Davis, CA: University of California, Davis, School of Education. Retrieved from: http://education.ucdavis.edu/sites/education.ucdavis.edu/files/ELP_Report.pdf

Agresti, A., & Finelay, B. (1997). *Statistical methods for the social sciences.* Englewood Cliffs, NJ: Prentice-Hall.

Allison, P.D. (2002). *Missing Data*, Thousand Oaks, CA: Sage Publications.

Ardasheva, Y., Tretter, T.R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the threshold hypothesis. *Language Learning 62*(3), 769-812. doi:10.1111/j.1467-9922.2011.00652.x

Arizona Secretary of State. (2000). *Proposition 203.* Retrieved from http://www.azsos.gov/election/2000/info/pubpamphlet/english/prop203.htm

August, D., & Hakuta, K. (1997). *Improving schooling for language minority children: A research agenda.* Washington, DC: National Research Council, Institute of Medicine.

August, D. & Shanahan, T., eds. (2006). *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum.

Bilingual Education Act, Pub. L. No. (90-247), 81 Stat. 816 (1968).

Calderon, M., Slavin, R., & Sanchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103-127. doi:10.1353/foc.2011.0007

California Department of Education. (1999). *Proposition 227 survey, interim report*. Sacramento: CDE.

Callahan, R. (2006). Intersection of Liability and Language: Reading Intervention vs. English Language Development. *Bilingual Research Journal, 30* (1), 1-21.

Cawthon, S. (2010). Assessment accommodations for students who are English Language Learners: The case of former-LEPs. *Practical Assessment, Evaluation, and Research, 15*(13). Retrieved from http://pareonline.net/pdf/v15n13.pdf

Cheema, J.R. (2014). A review of missing data handling methods in education research. *Review of Educational Research, 84*(4), 487-508.

Civil Rights Act, Pub. L. No. (88-352), 78 Stat. 255 (1964).

Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Collier, V. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly, 21,* 617-641.

Collier, V. (1995). *Promoting academic success for ESL students: Understanding second language acquisition for school.* Elizabeth, NJ: New Jersey Teachers of English to Speakers of Other Languages-Bilingual Educators.

Cox, B.E., McIntosh, K., Reason, R.D., Terenzini, P.T. (2014). Working with missing data in higher education research: A primer and real-world example. *The Review of Higher Education, 37*(3), 377-402.

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism, 19,* 121-129.

Cummins, J. (1981). The role of primary language development in promoting success for language minority students. In California Department of Education (Ed.), *Schooling and Language Minority Students: A Theoretical Framework*, 3-49. Los Angeles, CA.

Cummins, J. (2001). Empowering minority students: A framework for intervention. *Harvard Educational Review, 71*(4), 656-675.

Curtin, E.A. (2005). Teaching practices for ESL students. *Multicultural Education, 12*(3), 22-27.

Davies, S., O'Malley, K., & Wu, B. (2007, April). *Establishing measurement equivalence of transadapted reading and mathematics tests.* Paper presented at the annual meeting of American Educational Research Association, Chicago.

de Jong, E.J. (2004). After exit: Academic achievement patterns of former English language learners. *Education Policy Analysis Archives, 12*(50).

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*: 1-38.

Diggle, P.J., Liang, K., & Zeger, S. (1994). *Analysis of longitudinal data.* New York: Oxford University Press.

Echevarria, J., Vogt, M.E., & Short, D. (2000). *Making content comprehensible for English language learners: The SIOP model*. Needham Heights, MA: Allyn & Bacon.

Echevarria, J., Short, D., & Powers, K. (2006). School reform and standards-based education: A model for English-language learners. *Journal of Educational Research*, *99*(4), 195-210.

Elementary and Secondary Education Act, Pub. L. No. (89-10), 79 Stat. 27 (1965).

Equal Educational Opportunity Act, Pub, L. No. (93-380), 88 Stat. 515 (1974).

Flores, S.M., Batalova, J., & Fix, M. (2012). *The Educational Trajectories of English-Language Learners in Texas.* Washington, DC: Migration Policy Institute.

Florida, R., Mellander, C., (2015). *Segregated City: The Geography of Economic Segregation in America's Metros.* Toronto, ON: Martin Prosperity Institute.

Fry, R. & Gonzales, F. (2008, August) *One-in-Five and Growing Fast: A Profile of Hispanic Public School Students.* Washington, DC: Pew Hispanic Center.

Galvan-Luis, S. (2010). *The impact of bilingual education on academic achievement and language development of third grade English language learners in Texas.* Retrieved from ProQuest Dissertations and Theses. (UMI No. 3467984, Texas A&M University - Corpus Christi).

Garcia, E. E. (1991). *The education of linguistically and culturally diverse students: Effective instructional practices*. Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.

Genesee, F. (Ed.) (1999). *Program alternatives for linguistically diverse students.* Educational Practice Report 1. Berkeley, CA: Center for Research on Education, Diversity & Excellence.

Genesee, F., Lindholm-Leary, K. Saunders, W., and Christian, D. (2006). *Educating English Language Learners*. New York: Cambridge University Press.

Goldenberg, C. (2008). Teaching English language learners: What the research does—and does not—say. *American Educator, 32*(2), 8–44.

Guerrero, M.D. (2004). Acquiring academic English in one year: An unlikely proposition for English language learners. *Urban Education, 39*(2), 172-199. doi:10.1177/0042085903260915

Halle, T., Hair, E., Wandner, L. McNamara, M., & Chien, N. (2012). Predictors and outcomes of early versus later English language proficiency among English language learners. *Early Childhood Research Quarterly, 27*(1). 1-20. doi:10.1016/j.ecresq.2011.07.004

Hempel, L.M., Dowling, J.A., Boardman, J.D., & Ellison, C.G. (2013). Racial threat and white opposition to bilingual education in Texas. *Hispanic Journal of Behavioral Sciences, 35*(1), 85-102.

Hornberger, N. H. (1990). Bilingual education and English-Only: A language-planning framework. *The Annals of the American Academy of Political and Social Science*, *508*(1), 12-26.

Horne v. Flores, 557 U.S. 433 (LexisNexis 2009).

Huntington, S.P. (2004, March/April). The Hispanic challenge. *Foreign Policy,* 30-45.

Hursh, D. (2005). The growth of high-stakes testing in the USA: accountability, markets and the decline in educational equality. *British Educational Research Journal, 31*(5), 605-622.

Iddings, A.C.D. & Katz, L. (2007). Integrating home and school identities of recent-immigrant Hispanic English language learners through classroom practices. *Journal of Language, Identity, and Education, 6*(4), 299-314.

Kim, J., & Herman, J.L. (2009). *A three-state study of English learner progress* (CRESST Report 764). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Kim, J., & Herman, J.L. (2012). *Understanding patterns and precursors of ELL success subsequent to reclassification* (CRESST Report 818). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Kinder, D.R., & Sears, D.O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology, 40*(3), 414-431.

Krashen, S. (2004). *The acquisition of academic English by children in two-way programs: What does research say?* Paper presented at the National Association for Bilingual Education conference, Albuquerque, NM. Retrieved from http://www.sdkrashen.com

Ladson-Billings, G., & Tate, W. (1995). Toward a critical race theory of education. *Teachers College Record, 97*(1), 47-68.

Lara-Alecio, R., Galloway, M., Irby, B., Rodriquez, L., Gomez, L. (2004). Two-way immersion bilingual programs in Texas. *Bilingual Research Journal, 28*(1), 35-54.

Lau v. Nichols, 414 U.S. 563 (LexisNexis 1974).

Lindsey, J.K. (1993). *Models for repeated measures.* New York: Oxford University Press.

Losen, D., Orfield, G., & Balfanz, R. (2006). *Confronting the Graduation Rate Crisis in Texas.* Cambridge MA: The Civil Rights Project at Harvard University.

Lubienski, S.T. (2002). A closer look at Black-White mathematics gaps: Intersections of race and SES in NAEP achievement and instructional practices data. *Journal of Negro Education, 71*(4), 269-287.

Massachusetts Department of Education. (2002). *Questions and answers regarding 71A: English language education in public schools.* Retrieved from http://www.doe.mass.edu/ell/chapter71A_faq.pdf

Menken, K. & Kleyn, T. (2009). The Difficult Road for Long-Term English Learners. *Educational Leadership* (online only), *66(7).* Retrieved from http://www.ascd.org/publications/educational_leadership/apr09/vol66/num07/The_Difficult_Road_for_Long-Term_English_Learners.aspx

Mora, J.K., Wink, J., & Wink, D. (2001). Dueling models of dual language instruction: A critical review of the literature and program implementation guide. *Bilingual Research Journal, 25*(4), 101-110.

McNeil, L.M., Coppola, E., Radigan, J., & Vasquez-Heilig, J. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives, 16*(3). Retrieved from http://epaa.asu.edu/epaa/v16n3/

Nakamoto, J., Lindsey, K.A., & Manis, F.R. (2012). Development of reading skills from K-3 in Spanish-speaking English language learners following three programs of instruction. *Reading And Writing, 25*(2), 537-567. doi:10.1007/s11145-010-9285-4

National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs. (2007). *Texas: Rate of LEP growth, 1995/96 - 2005/06*. Author.

National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs (2009). *Glossary of terms related to the education of linguistically and culturally diverse students.* Author. Retrieved from: http://www.ncela.gwu.edu/files/rcd/BE021775/Glossary_of_Terms.pdf

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Olsen, L. (2010). *Reparable harm: Fulfilling the unkept promise of educational opportunity for long-term English learners*. Long Beach, California: Californians Together.

Orozco, M. (2012). *The effect of a pre-kindergarten early intervention program on the academic achievement of third-grade Hispanic English language learners.* Retrieved from ProQuest Dissertations and Theses. (UMI No. 3515692, University of Houston).

Peal, E., & Lambert, W.E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs, 76,* 1-23.

Peugh, J.L., & Enders, C.K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525-556

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousands Oaks, CA: Sage Publications.

Raudenbush, S.W., & Liu, X.F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods* 6(4): 387-401.

Roebuck, J. (2010). Court-ordered ESL overhaul scrapped on appeal. *The Monitor,* March 23, 2010. Retrieved from: http://www.themonitor.com/article_f6eafb8f-8df4-57ef-a9c1-be7648f7b1c1.html

Rolstad, K., Mahoney, K., & Glass, G.V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational policy, 19*(4), 572-594.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Survey*. New York: Wiley.

Ruiz, R. (1984). Orientations in language planning. *NABE Journal,* 7, 15-34.

San Miguel, G., Jr. (2004). *Contested policy: The rise and fall of federal bilingual education in the United States, 1960-2001.* Denton: University of North Texas Press.

Schafer, J.L. (1999). Multiple imputation: a primer. *Statistical Methods in medical Research,* 8, 3-15.

Senate Committee on Health, Education, Labor and Pensions. (2015). *The Every Child Achieves Act of 2015.* Retrieved from: http://www.help.senate.gov/imo/media/S_EveryChildAchievesActof2015.pdf

Slama, R.B. (2012). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal Of Educational Psychology 104*(2), 265-285. doi:10.1037/a0025861

Slavin, R.E., Madden, N., Calderon, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation And Policy Analysis, 33*(1), 47-58. doi:10.3102/0162373711398127

Smith, C. (2010, February). *Understanding the drop in achievement for students with Limited English Proficiency from elementary to secondary school.* Paper presented at The 33rd Annual Meeting of Southwest Educational Research Association. New Orleans, LA.

Snyder, T., & Musu-Gillette, L. (2015). Free or reduced price lunch: A proxy for poverty? *National Center for Educational Statistics Blog,* April 16, 2015. Retrieved from: http://nces.ed.gov/blogs/nces/post/free-or-reduced-price-lunch-a-proxy-for-poverty

Spybrook, J., Randenbush, S.W., Cogdon, R., & Martinez, A. (2009). Optimal design software. Retrieved from http://sitemaker.umich.edu/group-based/optimal_design_software

Stevens, J. (2007). *Intermediate Statistics: A Modern Approach* (3rd ed.). New York: Lawrence Erlbaum Associates.

Stewner-Manzanares, G., & National Clearinghouse for Bilingual Education, W. (1988). The Bilingual Education Act: Twenty Years Later. New Focus, Occasional Papers in Bilingual Education, Number 6. *New Focus*, (6).

Swanson, C. (2008). Cities in crisis: A special analytic report on high school graduation. Bethesda, MD: Editorial Projects in Education Research Center.

Texas Administrative Code. (2008). Title 19, Part 2, Chapter 101, Subchapter AA, Rule 101.1007: Limited English proficient students at grades other than the exit level.

Texas Administrative Code. (2011). Title 19, Part 2, Chapter 101, Subchapter AA. Commissioner's rules concerning the participation of English language learners in state assessments.

Texas Administrative Code. (2012). Title 19, Part 2, Chapter 89, Subchapter BB. Commissioner's rules concerning state plan for educating English language learners.

Texas Education Agency. (2008). *Secondary school completion and dropouts in Texas public schools, 2006-07* (Document No. GE08 601 05). Austin, TX: Author.

Texas Education Agency. (2009). *Enrollment in Texas public schools, 2007-08.* (Document No. GE09 601 06). Austin TX: Author.

Texas Education Agency (2010). *Student Assessment Division: Technical Digest 2008-2009.* Retrieved from: http://www.tea.state.tx.us/index3.aspx?id=2147484418&menu_id=793

Texas Education Agency (2011). *District Type Data, 2009-10.* Retrieved from http://tea.texas.gov/acctres/analyze/0910/level.html

Texas Education Agency. (2012a). *Enrollment in Texas public schools, 2011-12.* (Document No. GE13 601 02). Austin TX: Author.

Texas Education Agency. (2012b). *LPAC framework manual 2012-2013*. Curriculum Division. Austin TX: Author.

Texas Education Agency. (2013). *TEA seeks waiver of NCLB provisions.* Retrieved from http://www.tea.state.tx.us/index2.aspx?id=25769803880

Texas Legislature Online. (2009). *Text of House Bill 3.* Retrieved from http://www.capitol.state.tx.us/tlodocs/81R/billtext/pdf/HB00003F.pdf

Thomas, W.P., & Collier, V. (1997). School effectiveness for language minority students. Washington, DC: NCBE.

Thomas, W.P., Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Final report. Santa Cruz, CA: Center for Research on Education, Diversity & Excellence.

U.S. Department of Education. (2006). *Public Elementary and Secondary Students, Staff, Schools, and School Districts: School Year 2003-04* (NCES 2006–307). National Center for Education Statistics.

U.S. Department of Education. (2009). *High school drop out and completion rates in the United States: 2007.* National Center for Educational Statistics, Institute of Educational Sciences. (NCES 2009064).

U.S. Department of Education (2010). *Average reading scale scores of 4th- and 8th-graders in public schools and percentage scoring at or above selected reading achievement levels, by English language learner (ELL) status and state or jurisdiction: 2009.* National Center for Education Statistics, National Assessment of Educational Progress.

U.S. Department of Education (2012a). *The Condition of Education 2012* (NCES 2012-045). National Center for Education Statistics.

U.S. Department of Education. (2012b). *The nation's report card: Mathematics highlights 2011.* National Center for Educational Statistics, Institute of Educational Sciences.

U.S. Department of Education. (2012c). *The nation's report card: Reading highlights 2011.* National Center for Educational Statistics, Institute of Educational Sciences.

U.S. Department of Education. (2013). *Evidence review protocol for interventions for English language learners. Version 2.2.* Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse.

U.S. Department of Education. (2015). *ESEA Flexibility.* Office of Elementary & Secondary Education. Retrieved from: http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

Valencia, R.R. (2008). *Chicano students and the courts: The Mexican American legal struggle for educational equality.* New York and London: New York University Press.

Vasquez-Heilig, J., Darling-Hammond L. (2008). Accountability Texas-Style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75-110. doi:10.3102/016237370831768